PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
A01N 43/04, 59/16, A61K 9/48, 31/70, 31/715, 39/02, 39/40, G01N 33/554, 33/569

A1

(11) International Publication Number: WO 98/43479

(43) International Publication Date: 8 October 1998 (08.10.98)

(21) International Application Number: PCT/US98/06421

(22) International Filing Date: 31 March 1998 (31.03.98)

(30) Priority Data:

08/831,310	1 April 1997 (01.04.97)	US
08/834,666	1 April 1997 (01.04.97)	US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Applications

US	08/831,310 (CON)
Filed on	1 April 1997 (01.04.97)
US	08/834,666 (CON)
Filed on	1 April 1997 (01.04.97)

(71) Applicants *(for all designated States except US)*: MERIEUX ORAVAX SOCIETE EN NOM COLLECTIF, PASTEUR MERIEUX SERUMS ET VACCINS S.A. [FR/FR]; 58, avenue Leclerc, F–69007 Lyon (FR). HUMAN GENOME SCIENCES, INC. [US/US]; 9410 Key West Avenue, Rockville, MD 20850 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: KLEANTHOUS, Harold [GB/US]; 89 Madison Avenue, Newtonville, MA 02160 (US). LISSOLO, Ling [FR/FR]; 691, rue du Vallor, F–69280 Marcy l'Etoile (FR). TOMB, Jean–François [–/US]; 3501 St. Paul Street, Baltimore, MD 21222 (US). MILLER, Charles [US/US]; 32 Maple Avenue, Medford, MA 02155 (US). AL–GARAWI, Amal [SA/US]; 32 Garrison Street, No. 4501, Boston, MA 02114 (US).

(74) Agent: CLARK, Paul, T.; Clark & Elbing LLP, 176 Federal Street, Boston, MA 02110–2214 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: 76 kDa, 32 kDa, AND 50 kDa *HELICOBACTER* POLYPEPTIDES AND CORRESPONDING POLYNUCLEOTIDE MOLECULES

(57) Abstract

The invention provides 76 kDa, 32 kDa, and 50 kDa *Helicobacter* polypeptides, which can be used in vaccination methods for preventing or treating *Helicobacter* infection, and polynucleotides that encode these polypeptides. The invention also provides diagnostic methods employing these polypeptides.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

76 kDa, 32 kDa, and 50 kDa *HELICOBACTER* POLYPEPTIDES AND CORRESPONDING POLYNUCLEOTIDE MOLECULES

The invention relates to *Helicobacter* polypeptides and corresponding polynucleotide molecules that can be used in methods to prevent or treat *Helicobacter* infection in mammals, such as humans.

Background of the Invention

Helicobacter is a genus of spiral, gram-negative bacteria that colonize the gastrointestinal tracts of mammals. Several species colonize the stomach, most notably *H. pylori, H. heilmanii, H. felis*, and *H. mustelae*. Although *H. pylori* is the species most commonly associated with human infection, *H. heilmanii* and *H. felis* have also been isolated from humans, but at lower frequencies than *H. pylori*. *Helicobacter* infects over 50% of adult populations in developed countries and nearly 100% in developing countries and some Pacific rim countries, making it one of the most prevalent infections worldwide.

Helicobacter is routinely recovered from gastric biopsies of humans with histological evidence of gastritis and peptic ulceration. Indeed, *H. pylori* is now recognized as an important pathogen of humans, in that the chronic gastritis it causes is a risk factor for the development of peptic ulcer diseases and gastric carcinoma. It is thus highly desirable to develop safe and effective vaccines for preventing and treating *Helicobacter* infection.

A number of *Helicobacter* antigens have been characterized or isolated. These include urease, which is composed of two structural subunits of approximately 30 and 67 kDa (Hu *et al.*, Infect. Immun. 58:992, 1990; Dunn *et al.*, J. Biol. Chem. 265:9464, 1990; Evans *et al.*, Microbial Pathogenesis 10:15,

1991; Labigne *et al.*, J. Bact., 173:1920, 1991); the 87 kDa vacuolar cytotoxin (VacA) (Cover *et al.*, J. Biol. Chem. 267:10570, 1992; Phadnis *et al.*, Infect. Immun. 62:1557, 1994; WO 93/18150); a 128 kDa immunodominant antigen associated with the cytotoxin (CagA, also called TagA; WO 93/18150; U.S. Patent No. 5,403,924); 13 and 58 kDa heat shock proteins HspA and HspB (Suerbaum *et al.*, Mol. Microbiol. 14:959, 1994; WO 93/18150); a 54 kDa catalase (Hazell *et al.*, J. Gen. Microbiol.137:57, 1991); a 15 kDa histidine-rich protein (Hpn) (Gilbert *et al.*, Infect. Immun. 63:2682, 1995); a 20 kDa membrane-associated lipoprotein (Kostrcynska *et al.*, J. Bact. 176:5938, 1994); a 30 kDa outer membrane protein (Bölin *et al.*, J. Clin. Microbiol. 33:381, 1995); a lactoferrin receptor (FR 2,724,936); and several porins, designated HopA, HopB, HopC, HopD, and HopE, which have molecular weights of 48-67 kDa (Exner *et al.*, Infect. Immun. 63:1567, 1995; Doig *et al.*, J. Bact. 177:5447, 1995). Some of these proteins have been proposed as potential vaccine antigens. In particular, urease is believed to be a vaccine candidate (WO 94/9823; WO 95/22987; WO 95/3824; Michetti *et al.*, Gastroenterology 107:1002, 1994). Nevertheless, it is thought that several antigens may ultimately be necessary in a vaccine.

Summary of the Invention

The invention provides polynucleotide molecules that encode a family of 76 kDa *Helicobacter* polypeptides, designated GHPO 386, GHPO 789, GHPO 1516, GHPO 1197, GHPO 1180, GHPO 896, GHPO 711, GHPO 190, GHPO 185, GHPO 1417, and GHPO 1414, a 32 kDa polypeptide, designated GHPO 1360, and a 50 kDa polypeptide, designated GHPO 750, which can be used, *e.g.*, in methods to prevent, treat, or diagnose *Helicobacter* infection. The polypeptides include those having the amino acid sequences

shown in SEQ ID NOs:2-22 (even numbers), 66, and 68. Those skilled in the art will understand that the invention also includes polynucleotide molecules that encode mutants and derivatives of these polypeptides, which can result from the addition, deletion, or substitution of non-essential amino acids, as is described further below.

In addition to the polynucleotide molecules described above, the invention includes the corresponding polypeptides (*i.e.*, polypeptides encoded by the polynucleotide molecules of the invention, or fragments thereof), and monospecific antibodies that specifically bind to these polypeptides.

The present invention has many applications and includes expression cassettes, vectors, and cells transformed or transfected with the polynucleotides of the invention. Accordingly, the present invention provides (i) methods for producing polypeptides of the invention in recombinant host systems and related expression cassettes, vectors, and transformed or transfected cells; (ii) live vaccine vectors, such as pox virus, *Salmonella typhimurium*, and *Vibrio cholerae* vectors, that contain polynucleotides of the invention (such vaccine vectors being useful in, *e.g.*, methods for preventing or treating *Helicobacter* infection) in combination with a diluent or carrier, and related pharmaceutical compositions and associated therapeutic and/or prophylactic methods; (iii) therapeutic and/or prophylactic methods involving administration of polynucleotide molecules, either in a naked form or formulated with a delivery vehicle, polypeptides or mixtures of polypeptides, or monospecific antibodies of the invention, and related pharmaceutical compositions; (iv) methods for detecting the presence of *Helicobacter* in biological samples, which can involve the use of polynucleotide molecules, monospecific antibodies, or polypeptides of the invention; and (v) methods for purifying polypeptides of the invention by antibody-based affinity chromatography.

Brief Description of the Drawings

Figure 1 is an alignment of the predicted amino acid sequences of GHPO 386 (SEQ ID NO:2), GHPO 789 (SEQ ID NO:4), and GHPO 1516 (SEQ ID NO:6), as well as a consensus sequence for the 76 kDa protein family.

Figure 2 is an alignment of the predicted amino acid sequences of GHPO 1197 (SEQ ID NO:8), GHPO 1180 (SEQ ID NO:10), GHPO 896 (SEQ ID NO:12), GHPO 711 (SEQ ID NO:14), GHPO 190 (SEQ ID NO:16), GHPO 185 (SEQ ID NO:18), GHPO 1417 (SEQ ID NO:20), and GHPO 1414 (SEQ ID NO:22), as well as a consensus sequence for the 76 kDa protein family.

Detailed Description

Open reading frames (ORFs) encoding a family of new, full length, membrane-associated 76 kDa polypeptides, designated GHPO 386, GHPO 789, GHPO 1516, GHPO 1197, GHPO 1180, GHPO 896, GHPO 711, GHPO 190, GHPO 185, GHPO 1417, and GHPO 1414, a 32 kDa polypeptide, designated GHPO 1360, and a 50 kDa polypeptide, designated GHPO 750, have been identified in the *H. pylori* genome. The amino acid sequences of the 76 kDa polypeptides are aligned in Figures 1 and 2. The 76 kDa, 32 kDa, and 50 kDa polypeptides can be used, for example, in vaccination methods for preventing or treating *Helicobacter* infection. For example, GHPO 750, GHPO 1360, GHPO 190, and GHPO 1516 have been shown to be protective antigens. By "protective antigen" is meant an antigen that is capable of reducing the infection level after challenge, relative to a positive control. Absolute protection from infection, although included in the invention, is not required.

The polypeptides of the invention (except GHPO 750, see below) are secreted polypeptides that can be produced in their mature forms (*i.e.*, as polypeptides that have been exported through class II or class III secretion

pathways) or as precursors that include a signal peptide, which can be removed in the course of excretion/secretion by cleavage at the N-terminal end of the mature form. (The cleavage site is located at the C-terminal end of the signal peptide, adjacent to the mature form.) The cleavage site for the polypeptides of the invention and, thus, the first amino acid of the mature polypeptides, was putatively determined.

According to a first aspect of the invention, there are provided isolated polynucleotides that encode the precursor and mature forms of *Helicobacter* GHPO 386, GHPO 789, GHPO 1516, GHPO 1197, GHPO 1180, GHPO 896, GHPO 711, GHPO 190, GHPO 185, GHPO 1417, GHPO 1414, GHPO 1360, and GHPO 750.

An isolated polynucleotide of the invention encodes:

(i) a polypeptide having an amino acid sequence that is homologous to a *Helicobacter* amino acid sequence of a polypeptide associated with the *Helicobacter* membrane, the *Helicobacter* amino acid sequence being selected from the group consisting of the amino acid sequences shown:

-in SEQ ID NO:2, beginning with an amino acid in any one of positions -19 to 5, preferably in position -19 or position 1, and ending with an amino acid in position 689 (GHPO 386);

-in SEQ ID NO:4, beginning with an amino acid in any one of positions -20 to 5, preferably in position -20 or position 1, and ending with an amino acid in position 713 (GHPO 789);

-in SEQ ID NO:6, beginning with an amino acid in any one of positions -20 to 5, preferably in position -20 or position 1, and ending with an amino acid in position 725 (GHPO 1516);

-in SEQ ID NO:8, beginning with an amino acid in any one of positions -20 to 5, preferably in position -20 or position 1, and ending with an amino acid in position 691 (GHPO 1197);

-in SEQ ID NO:10, beginning with an amino acid in any one of positions -20 to 5, preferably in position -20 or position 1, and ending with an amino acid in position 652 (GHPO 1180);

-in SEQ ID NO:12, beginning with an amino acid in any one of positions -18 to 5, preferably in position -18 or position 1, and ending with an amino acid in position 673 (GHPO 896);

-in SEQ ID NO:14, beginning with an amino acid in any one of positions -21 to 5, preferably in position -21 or position 1, and ending with an amino acid in position 619 (GHPO 711);

-in SEQ ID NO:16, beginning with an amino acid in any one of positions -17 to 5, preferably in position -17 or position 1, and ending with an amino acid in position 635 (GHPO 190);

-in SEQ ID NO:18, beginning with an amino acid in any one of positions -19 to 5, preferably in position -19 or position 1, and ending with an amino acid in position 626 (GHPO 185);

-in SEQ ID NO:20, beginning with an amino acid in any one of positions -16 to 5, preferably in position -16 or position 1, and ending with an amino acid in position 467 (GHPO 1417);

-in SEQ ID NO:22, beginning with an amino acid in any one of positions -18 to 5, preferably in position -18 or position 1, and ending with an amino acid in position 673 (GHPO 1414);

-in SEQ ID NO:66, beginning with an amino acid in any one of positions -20 to 5, preferably in position -20 or position 1, and ending with an amino acid in position 279 (GHPO 1360); and

-in SEQ ID NO:68, beginning with an amino acid in position 1 and ending with an amino acid in position 399 (GHPO 750); or

(ii) a derivative of the polypeptide.

The term "isolated polynucleotide" is defined as a polynucleotide that is removed from the environment in which it naturally occurs. For example, a naturally-occurring DNA molecule present in the genome of a living bacteria or as part of a gene bank is not isolated, but the same molecule, separated from the remaining part of the bacterial genome, as a result of, *e.g.*, a cloning event (amplification), is "isolated." Typically, an isolated DNA molecule is free from DNA regions (*e.g.*, coding regions) with which it is immediately contiguous, at the 5' or 3' ends, in the naturally occurring genome. Such isolated polynucleotides can be part of a vector or a composition and still be isolated, as such a vector or composition is not part of its natural environment.

A polynucleotide of the invention can consist of RNA or DNA (*e.g.*, cDNA, genomic DNA, or synthetic DNA), or modifications or combinations of RNA or DNA. The polynucleotide can be double-stranded or single-stranded and, if single-stranded, can be the coding (sense) strand or the non-coding (anti-sense) strand. The sequences that encode polypeptides of the invention, as shown in SEQ ID NOs:2-22 (even numbers), 66, and 68, can be (a) the coding sequence as shown in SEQ ID NOs:1-21 (odd numbers), 65, and 67; (b) a ribonucleotide sequence derived by transcription of (a); or (c) a different coding sequence that, as a result of the redundancy or degeneracy of the genetic code, encodes the same polypeptides as the polynucleotide molecules having the sequences illustrated in any of SEQ ID NOs:1-21 (odd numbers), 65, and 67. The polypeptides of the invention can be ones that are naturally secreted or excreted by, *e.g.*, *H. felis, H. mustelae, H. heilmanii*, or *H. pylori.*

By "polypeptide" or "protein" is meant any chain of amino acids, regardless of length or post-translational modification (*e.g.*, glycosylation or phosphorylation). Both terms are used interchangeably in the present application.

By "homologous amino acid sequence" is meant an amino acid sequence that differs from an amino acid sequence shown in any of SEQ ID NOs:2-22 (even numbers), 66, and 68, or an amino acid sequence encoded by the nucleotide sequence of any of SEQ ID NOs:1-21 (odd numbers), 65, and 67, by one or more non-conservative amino acid substitutions, deletions, or additions located at positions at which they do not destroy the specific antigenicity of the polypeptide. Preferably, such a sequence is at least 75%, more preferably at least 80%, and most preferably at least 90% identical to an amino acid sequence shown in any of SEQ ID NOs:2-22 (even numbers), 66, and 68.

Homologous amino acid sequences include sequences that are identical or substantially identical to an amino acid sequence as shown in any of SEQ ID NOs:2-22 (even numbers), 66, and 68. By "amino acid sequence that is substantially identical" is meant a sequence that is at least 90%, preferably at least 95%, more preferably at least 97%, and most preferably at least 99% identical to an amino acid sequence of reference and that differs from the sequence of reference, if at all, by a majority of conservative amino acid substitutions.

Conservative amino acid substitutions typically include substitutions among amino acids of the same class. These classes include, for example, amino acids having uncharged polar side chains, such as asparagine, glutamine, serine, threonine, and tyrosine; amino acids having basic side chains, such as lysine, arginine, and histidine; amino acids having acidic side chains, such as

aspartic acid and glutamic acid; and amino acids having nonpolar side chains, such as glycine, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan, and cysteine.

Homology can be measured using sequence analysis software (*e.g.*, Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, WI 53705). Similar amino acid sequences are aligned to obtain the maximum degree of homology (*i.e.*, identity). To this end, it may be necessary to artificially introduce gaps into the sequence. Once the optimal alignment has been set up, the degree of homology (*i.e.*, identity) is established by recording all of the positions in which the amino acids of both sequences are identical, relative to the total number of positions.

Homologous polynucleotide sequences are defined in a similar way. Preferably, a homologous sequence is one that is at least 45%, more preferably at least 60%, and most preferably at least 85% identical to a coding sequence of any of SEQ ID NOs:1-21 (odd numbers), 65, and 67.

Polypeptides having a sequence homologous to one of the sequences shown in SEQ ID NOs:2-22 (even numbers), 66, and 68 include naturally-occurring allelic variants, as well as mutants or any other non-naturally occurring variants that are analogous in terms of antigenicity, to a polypeptide having a sequence as shown in SEQ ID NOs:2-22 (even numbers), 66, and 68.

As is known in the art, an allelic variant is an alternate form of a polypeptide that is characterized as having a substitution, deletion, or addition of one or more amino acids that does not alter the biological function of the polypeptide. By "biological function" is meant a function of the polypeptide in the cells in which it naturally occurs, even if the function is not necessary for the growth or survival of the cells. For example, the biological function of a

porin is to allow the entry into cells of compounds present in the extracellular medium. The biological function is distinct from the antigenic function. A polypeptide can have more than one biological function.

Allelic variants are very common in nature. For example, a bacterial species, *e.g.*, *H. pylori,* is usually represented by a variety of strains that differ from each other by minor allelic variations. Indeed, a polypeptide that fulfills the same biological function in different strains can have an amino acid sequence that is not identical in each of the strains. Such an allelic variation can be equally reflected at the polynucleotide level.

Support for the use of allelic variants of polypeptide antigens comes from, *e.g.*, studies of the *Helicobacter* urease antigen. The amino acid sequence of *Helicobacter* urease varies widely from species to species, yet cross-species protection occurs, indicating that the urease molecule, when used as an immunogen, is highly tolerant of amino acid variations. Even among different strains of the single species *H. pylori*, there are amino acid sequence variations.

For example, although the amino acid sequences of the UreA and UreB subunits of *H. pylori* and *H. felis* ureases differ from one another by 26.5% and 11.8%, respectively (Ferrero *et al.*, Molecular Microbiology 9(2):323-333, 1993), it has been shown that *H. pylori* urease protects mice from *H. felis* infection (Michetti *et al.*, Gastroenterology 107:1002, 1994). In addition, it has been shown that the individual structural subunits of urease, UreA and UreB, which contain distinct amino acid sequences, are both protective antigens against *Helicobacter* infection (Michetti *et al.*, *supra*). Similarly, Cuenca *et al.* (Gastroenterology 110:1770, 1996) showed that therapeutic immunization of *H. mustelae*-infected ferrets with *H. pylori* urease was effective at eradicating *H. mustelae* infection. Further, several urease

variants have been reported to be effective vaccine antigens, including, *e.g.*, recombinant UreA + UreB apoenzyme expressed from pORV142 (UreA and UreB sequences derived from *H. pylori* strain CPM630; Lee *et al.*, J. Infect. Dis.172:161, 1995); recombinant UreA + UreB apoenzyme expressed from pORV214 (UreA and UreB sequences differ from *H. pylori* strain CPM630 by one and two amino acid changes, respectively; Lee *et al.*, *supra*, 1995); a UreA-glutathione-S-transferase fusion protein (UreA sequence from *H. pylori* strain ATCC 43504; Thomas *et al.*, Acta Gastro-Enterologica Belgica 56:54, 1993); UreA + UreB holoenzyme purified from *H. pylori* strain NCTC11637 (Marchetti *et al.*, Science 267:1655, 1995); a UreA-MBP fusion protein (UreA from *H. pylori* strain 85P; Ferrero *et al.*, Infection and Immunity 62:4981, 1994); a UreB-MBP fusion protein (UreB from *H. pylori* strain 85P; Ferrero *et al.*, *supra*); a UreA-MBP fusion protein (UreA from *H. felis* strain ATCC 49179; Ferrero *et al.*, *supra*); a UreB-MBP fusion protein (UreB from *H. felis* strain ATCC 49179; Ferrero *et al.*, *supra*); and a 37 kDa fragment of UreB containing amino acids 220-569 (Dore-Davin *et al.*, "A 37 kD fragment of UreB is sufficient to confer protection against *Helicobacter felis* infection in mice"). Finally, Thomas *et al.* (*supra*) showed that oral immunization of mice with crude sonicates of *H. pylori* protected mice from subsequent challenge with *H. felis*.

Polynucleotides, *e.g.*, DNA molecules, encoding allelic variants can easily be obtained by polymerase chain reaction (PCR) amplification of genomic bacterial DNA extracted by conventional methods. This involves the use of synthetic oligonucleotide primers matching sequences that are upstream and downstream of the 5' and 3' ends of the coding region. Suitable primers can be designed based on the nucleotide sequence information provided in SEQ ID NOs:1-21 (odd numbers), 65, and 67. Typically, a primer consists of 10 to

40, preferably 15 to 25 nucleotides. It can also be advantageous to select primers containing C and G nucleotides in proportions sufficient to ensure efficient hybridization, *e.g.*, an amount of C and G nucleotides of at least 40%, preferably 50%, of the total nucleotide amount. Those skilled in the art can readily design primers that can be used to isolate the polynucleotides of the invention from different *Helicobacter* strains.

As an example, primers useful for cloning a polynucleotide molecule encoding a polypeptide having the amino acid sequence of unprocessed GHPO 386 (SEQ ID NO:2), including a signal peptide, are shown in SEQ ID NO:23 (matching at the 5' end) and in SEQ ID NO:25 (matching at the 3' end). Primers useful for cloning a DNA molecule encoding a polypeptide having the amino acid sequence of mature GHPO 386 (amino acids 1-689 of SEQ ID NO:2), lacking a signal peptide, are shown in SEQ ID NO:24 (matching at the 5' end) and in SEQ ID NO:25 (matching at the 3' end). Primers useful for cloning a DNA molecule encoding a polypeptide having the amino acid sequence of GHPO 1360 (SEQ ID NO:66), are shown in SEQ ID NO:78 (matching at the 5' end) and in SEQ ID NO:79 (matching at the 3' end). Use of these primers enables amplification of the entire gene encoding GHPO 1360. Primers having sequences shown in SEQ ID NO:82 (matching at the 5' end of the coding sequence corresponding to the mature protein) and SEQ ID NO:79 (matching at the 3' end) can be used to amplify the portion of the gene encoding mature GHPO 1360. Experimental conditions for carrying out PCR can readily be determined by one skilled in the art and illustrations of carrying out PCR are provided in Examples 3 and 4.

Thus, the first aspect of the invention includes:

(i) isolated polynucleotide molecules (*e.g.*, DNA molecules) that can be amplified and/or cloned using the polymerase chain reaction from a *Helicobacter*, *e.g.*, *H. pylori*, genome using either:

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:23, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:25 (unprocessed GHPO 386);

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:26, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:28 (unprocessed GHPO 789);

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:29, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:31 (unprocessed GHPO 1516);

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:32, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:34 (unprocessed GHPO 1197);

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:35, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:37 (unprocessed GHPO 1180);

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:38, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:40 (unprocessed GHPO 896);

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:41, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:43 (unprocessed GHPO 711);

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:44, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:46 (unprocessed GHPO 190);

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:47, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:49 (unprocessed GHPO 185);

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:50, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:52 (unprocessed GHPO 1417);

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:53, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:55 (unprocessed GHPO 1414);

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:78, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:79 (unprocessed GHPO 1360); or

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:80, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:81 (GHPO 750); and

(ii) isolated polynucleotide molecules (*e.g.*, DNA molecules) that can be amplified and/or cloned by the polymerase chain reaction from a *Helicobacter*, *e.g.*, *H. pylori*, genome using either:

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:24, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:25 (mature GHPO 386);

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:27, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:28 (mature GHPO 789);

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:30, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:31 (mature GHPO 1516);

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:33, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:34 (mature GHPO 1197);

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:36, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:37 (mature GHPO 1180);

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:39, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:40 (mature GHPO 896);

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:42, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:43 (mature GHPO 711);

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:45, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:46 (mature GHPO 190);

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:48, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:49 (mature GHPO 185);

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:51, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:52 (mature GHPO 1417);

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:54, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:55 (mature GHPO 1414); or

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:82, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:79 (mature GHPO 1360).

The 5' ends of the primers described above can advantageously include a restriction endonuclease recognition site that contains, typically, 4 to 6 nucleotides. For example, the sequences 5'-GGATCC-3' (*BamHI*) or 5'-CTCGAG-3' (*XhoI*) can be used. Restriction sites can be selected by those skilled in the art so that the amplified DNA, when digested, if necessary, can be conveniently cloned into an appropriately digested vector, such as a plasmid vector. In addition, a 5' clamp (*e.g.*, GCC) can be included in the primers 5' to the restriction endonuclease recognition site.

Useful homologs that do not occur naturally can be designed using known methods for identifying regions of an antigen that are likely to be tolerant of amino acid sequence changes and/or deletions. For example, sequences of the antigen from different species can be compared to identify conserved sequences.

Polypeptide derivatives that are encoded by polynucleotides of the invention include, *e.g.*, fragments, polypeptides having large internal deletions derived from full-length polypeptides, and fusion proteins. Polypeptide fragments of the invention can be derived from a polypeptide having a sequence homologous to the sequences of any of SEQ ID NOs:2-22 (even numbers), 66, and 68, to the extent that the fragments retain the substantial antigenicity of the parent polypeptide (specific antigenicity). Polypeptide derivatives can also be constructed by large internal deletions that remove a substantial part of the parent polypeptide, while retaining specific antigenicity. Generally, polypeptide derivatives should be about at least 12 amino acids in length to maintain antigenicity. Advantageously, they can be at least 20 amino acids, preferably at least 50 amino acids, more preferably at least 75 amino acids, and most preferably at least 100 amino acids in length.

Useful polypeptide derivatives, *e.g.*, polypeptide fragments, can be designed using computer-assisted analysis of amino acid sequences in order to identify sites in protein antigens having potential as surface-exposed, antigenic regions (Hughes *et al.*, Infect. Immun. 60(9):3497, 1992). For example, the Laser Gene Program from DNA Star can be used to obtain hydrophilicity, antigenic index, and intensity index plots for the polypeptides of the invention. This program can also be used to obtain information about homologies of the polypeptides with known protein motifs. One skilled in the art can readily use the information provided in such plots to select peptide fragments for use as vaccine antigens. For example, fragments spanning regions of the plots in which the antigenic index is relatively high can be selected. One can also select fragments spanning regions in which both the antigenic index and the intensity plots are relatively high. Fragments containing conserved sequences, particularly hydrophilic conserved sequences, can also be selected.

Polypeptide fragments and polypeptides having large internal deletions can be used for revealing epitopes that are otherwise masked in the parent polypeptide and that may be of importance for inducing a protective T cell-dependent immune response. Deletions can also remove immunodominant regions of high variability among strains.

It is an accepted practice in the field of immunology to use fragments and variants of protein immunogens as vaccines, as all that is required to induce an immune response to a protein is a small (*e.g.*, 8 to 10 amino acids) immunogenic region of the protein. This has been done for a number of vaccines against pathogens other than *Helicobacter*. For example, short synthetic peptides corresponding to surface-exposed antigens of pathogens such as murine mammary tumor virus (peptide containing 11 amino acids; Dion *et al.*, Virology 179:474-477, 1990), Semliki Forest virus (peptide containing 16

amino acids; Snijders *et al.*, J. Gen. Virol. 72:557-565, 1991), and canine parvovirus (2 overlapping peptides, each containing 15 amino acids; Langeveld *et al.*, Vaccine 12(15):1473-1480, 1994) have been shown to be effective vaccine antigens against their respective pathogens.

Polynucleotides encoding polypeptide fragments and polypeptides having large internal deletions can be constructed using standard methods (see, *e.g.*, Ausubel *et al.*, *Current Protocols in Molecular Biology*, John Wiley & Sons Inc., 1994), for example, by PCR, including inverse PCR, by restriction enzyme treatment of the cloned DNA molecules, or by the method of Kunkel *et al.* (Proc. Natl. Acad. Sci. USA 82:448, 1985; biological material available at Stratagene).

A polypeptide derivative can also be produced as a fusion polypeptide that contains a polypeptide or a polypeptide derivative of the invention fused, *e.g.*, at the – or C-terminal end, to any other polypeptide (hereinafter referred to as a peptide tail). Such a product can be easily obtained by translation of a genetic fusion, *i.e.*, a hybrid gene. Vectors for expressing fusion polypeptides are commercially available, and include the pMal-c2 or pMal-p2 systems of New England Biolabs, in which the peptide tail is a maltose binding protein, the glutathione-S-transferase system of Pharmacia, or the His-Tag system available from Novagen. These and other expression systems provide convenient means for further purification of polypeptides and derivatives of the invention.

Another particular example of fusion polypeptides included in invention includes a polypeptide or polypeptide derivative of the invention fused to a polypeptide having adjuvant activity, such as, *e.g.*, subunit B of either cholera toxin or *E. coli* heat-labile toxin. Several possibilities can be

used for producing such fusion proteins. First, the polypeptide of the invention can be fused to the

N-terminal end or, preferably, to the C-terminal end of the polypeptide having adjuvant activity. Second, a polypeptide fragment of the invention can be fused within the amino acid sequence of the polypeptide having adjuvant activity. Spacer sequences can also be included, if desired.

As stated above, the polynucleotides of the invention encode *Helicobacter* polypeptides in precursor or mature form. They can also encode hybrid precursors containing heterologous signal peptides, which can mature into polypeptides of the invention. By "heterologous signal peptide" is meant a signal peptide that is not found in the naturally-occurring precursor of a polypeptide of the invention.

A polynucleotide of the invention hybridizes, preferably under stringent conditions, to a polynucleotide having a sequence as shown in any of SEQ ID NOs:1-21 (odd numbers), 65, and 67. Hybridization procedures are, *e.g.*, described by Ausubel *et al.* (*supra*); Silhavy *et al.* (*Experiments with Gene Fusions*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1984); and Davis *et al. (A Manual for Genetic Engineering: Advanced Bacterial Genetics*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1980). Important parameters that can be considered for optimizing hybridization conditions are reflected in the following formula, which facilitates calculation of the melting temperature (Tm), which is the temperature above which two complementary DNA strands separate from one another (Casey *et al.*, Nucl. Acid Res. 4:1539, 1977): Tm = 81.5 + 0.5 x (% G+C) + 1.6 log (positive ion concentration) - 0.6 x (% formamide). Under appropriate stringency conditions, hybridization temperature (Th) is approximately 20 to 40°C, 20 to 25°C, or, preferably, 30 to 40°C below the

calculated Tm. Those skilled in the art will understand that optimal temperature and salt conditions can be readily determined empirically in preliminary experiments using conventional procedures. For example, stringent conditions can be achieved, both for pre-hybridizing and hybridizing incubations, (i) within 4-16 hours at 42°C, in 6 x SSC containing 50% formamide or (ii) within 4-16 hours at 65°C in an aqueous 6 x SSC solution (1 M NaCl, 0.1 M sodium citrate (pH 7.0)). For polynucleotides containing 30 to 600 nucleotides, the above formula is used and then is corrected by subtracting (600/polynucleotide size in base pairs). Stringency conditions are defined by a Th that is 5 to 10°C below Tm.

Hybridization conditions with oligonucleotides shorter than 20-30 bases do not precisely follow the rules set forth above. In such cases, the formula for calculating the Tm is as follows: Tm = 4 x (G+C) + 2 (A+T). For example, an 18 nucleotide fragment of 50% G+C would have an approximate Tm of 54°C.

A polynucleotide molecule of the invention, containing RNA, DNA, or modifications or combinations thereof, can have various applications. For example, a polynucleotide molecule can be used (i) in a process for producing the encoded polypeptide in a recombinant host system, (ii) in the construction of vaccine vectors such as poxviruses, which are further used in methods and compositions for preventing and/or treating *Helicobacter* infection, (iii) as a vaccine agent, in a naked form or formulated with a delivery vehicle, and (iv) in the construction of attenuated *Helicobacter* strains that can over-express a polynucleotide of the invention or express it in a non-toxic, mutated form.

According to a second aspect of the invention, there is therefore provided (i) an expression cassette containing a polynucleotide molecule of the invention placed under the control of elements (*e.g.*, a promoter) required for

expression; (ii) an expression vector containing an expression cassette of the invention; (iii) a procaryotic or eucaryotic cell transformed or transfected with an expression cassette and/or vector of the invention; as well as (iv) a process for producing a polypeptide or polypeptide derivative encoded by a polynucleotide of the invention, which involves culturing a procaryotic or eucaryotic cell transformed or transfected with an expression cassette and/or vector of the invention, under conditions that allow expression of the polynucleotide molecule of the invention and, recovering the encoded polypeptide or polypeptide derivative from the cell culture.

A recombinant expression system can be selected from procaryotic and eucaryotic hosts. Eucaryotic hosts include, for example, yeast cells (*e.g.*, *Saccharomyces cerevisiae* or *Pichia pastoris*), mammalian cells (*e.g.*, COS1, NIH3T3, or JEG3 cells), arthropods cells (*e.g.*, *Spodoptera frugiperda* (SF9) cells), and plant cells. Preferably, a procaryotic host such as *E. coli* is used. Bacterial and eucaryotic cells are available from a number of different sources that are known to those skilled in the art, *e.g.*, the American Type Culture Collection (ATCC; Rockville, Maryland).

The choice of the expression cassette will depend on the host system selected, as well as the features desired for the expressed polypeptide. For example, it may be useful to produce a polypeptide of the invention in a particular lipidated form or any other form. Typically, an expression cassette includes a constitutive or inducible promoter that is functional in the selected host system; a ribosome binding site; a start codon (ATG); if necessary, a region encoding a signal peptide, *e.g.*, a lipidation signal peptide; a polynucleotide molecule of the invention; a stop codon; and, optionally, a 3' terminal region (translation and/or transcription terminator). The signal peptide-encoding region is adjacent to the polynucleotide of the invention and

is placed in the proper reading frame. The signal peptide-encoding region can be homologous or heterologous to the polynucleotide molecule encoding the mature polypeptide and it can be specific to the secretion apparatus of the host used for expression. The open reading frame constituted by the polynucleotide molecule of the invention, alone or together with the signal peptide, is placed under the control of the promoter so that transcription and translation occur in the host system. Promoters and signal peptide-encoding regions are widely known and available to those skilled in the art and include, for example, the promoter of *Salmonella typhimurium* (and derivatives) that is inducible by arabinose (promoter araB) and is functional in Gram-negative bacteria such as *E. coli* (U.S. Patent No. 5,028,530; Cagnon *et al.*, Protein Engineering 4(7):843, 1991); the promoter of the bacteriophage T7 RNA polymerase gene, which is functional in a number of *E. coli* strains expressing T7 polymerase (U.S. Patent No. 4,952,496); the OspA lipidation signal peptide; and RlpB lipidation signal peptide (Takase *et al.*, J. Bact. 169:5692, 1987).

The expression cassette is typically part of an expression vector, which is selected for its ability to replicate in the chosen expression system. Expression vectors (*e.g.*, plasmids or viral vectors) can be chosen from, for example, those described in Pouwels *et al.* (*Cloning Vectors: A Laboratory Manual*, 1985, Supp. 1987) and can purchased from various commercial sources. Methods for transforming or transfecting host cells with expression vectors are well known in the art and will depend on the host system selected, as described in Ausubel *et al.* (*supra*).

Upon expression, a recombinant polypeptide of the invention (or a polypeptide derivative) is produced and remains in the intracellular compartment, is secreted/excreted in the extracellular medium or in the periplasmic space, or is embedded in the cellular membrane. The polypeptide

can then be recovered in a substantially purified form from the cell extract or from the supernatant after centrifugation of the cell culture. Typically, the recombinant polypeptide can be purified by antibody-based affinity purification or by any other method known to a person skilled in the art, such as by genetic fusion to a small affinity-binding domain. Antibody-based affinity purification methods are also available for purifying a polypeptide of the invention extracted from a *Helicobacter* strain. Antibodies useful for immunoaffinity purification of the polypeptides of the invention can be obtained using methods described below.

Polynucleotides of the invention can also be used in DNA vaccination methods, using either a viral or bacterial host as gene delivery vehicle (live vaccine vector) or administering the gene in a free form, *e.g.*, inserted into a plasmid. Therapeutic or prophylactic efficacy of a polynucleotide of the invention can be evaluated as is described below.

Accordingly, in a third aspect of the invention, there is provided (i) a vaccine vector such as a poxvirus, containing a polynucleotide molecule of the invention placed under the control of elements required for expression; (ii) a composition of matter containing a vaccine vector of the invention, together with a diluent or carrier; (iii) a pharmaceutical composition containing a therapeutically or prophylactically effective amount of a vaccine vector of the invention; (iv) a method for inducing an immune response against *Helicobacter* in a mammal (*e.g.*, a human; alternatively, the method can be used in veterinary applications for treating or preventing *Helicobacter* infection of animals, *e.g.*, cats or birds), which involves administering to the mammal an immunogenically effective amount of a vaccine vector of the invention to elicit an immune response, *e.g.*, a protective or therapeutic immune response to *Helicobacter*; and (v) a method for preventing and/or treating a *Helicobacter*

(*e.g., H. pylori, H. felis, H. mustelae*, or *H. heilmanii*) infection, which involves administering a prophylactic or therapeutic amount of a vaccine vector of the invention to an individual in need. Additionally, the third aspect of the invention encompasses the use of a vaccine vector of the invention in the preparation of a medicament for preventing and/or treating *Helicobacter* infection.

A vaccine vector of the invention can express one or several polypeptides or derivatives of the invention, as well as at least one additional *Helicobacter* antigen such as a urease apoenzyme or a subunit, fragment, homolog, mutant, or derivative thereof. In addition, it can express a cytokine, such as interleukin-2 (IL-2) or interleukin-12 (IL-12), that enhances the immune response. Thus, a vaccine vector can include an additional polynucleotide molecules encoding, *e.g.*, urease subunit A, B, or both, or a cytokine, placed under the control of elements required for expression in a mammalian cell.

Alternatively, a composition of the invention can include several vaccine vectors, each of which being capable of expressing a polypeptide or derivative of the invention. A composition can also contain a vaccine vector capable of expressing an additional *Helicobacter* antigen such as urease apoenzyme, a subunit, fragment, homolog, mutant, or derivative thereof, or a cytokine such as IL-2 or IL-12.

In vaccination methods for treating or preventing infection in a mammal, a vaccine vector of the invention can be administered by any conventional route in use in the vaccine field, for example, to a mucosal (*e.g.*, ocular, intranasal, oral, gastric, pulmonary, intestinal, rectal, vaginal, or urinary tract) surface or *via* a parenteral (*e.g.*, subcutaneous, intradermal, intramuscular, intravenous, or intraperitoneal) route. Preferred routes depend

upon the choice of the vaccine vector. The administration can be achieved in a single dose or repeated at intervals. The appropriate dosage depends on various parameters that are understood by those skilled in the art, such as the nature of the vaccine vector itself, the route of administration, and the condition of the mammal to be vaccinated (*e.g.*, the weight, age, and general health of the mammal).

Live vaccine vectors that can be used in the invention include viral vectors, such as adenoviruses and poxviruses, as well as bacterial vectors, *e.g.*, *Shigella, Salmonella*, *Vibrio cholerae*, *Lactobacillus*, Bacille bilié de Calmette-Guérin (BCG), and *Streptococcus*. An example of an adenovirus vector, as well as a method for constructing an adenovirus vector capable of expressing a polynucleotide molecule of the invention, is described in U.S. Patent No. 4,920,209. Poxvirus vectors that can be used in the invention include, *e.g.*, vaccinia and canary pox viruses, which are described in U.S. Patent No. 4,722,848 and U.S. Patent No. 5,364,773, respectively (also see, *e.g.*, Tartaglia *et al.*, Virology 188:217, 1992, for a description of a vaccinia virus vector, and Taylor *et al*, Vaccine 13:539, 1995, for a description of a canary poxvirus vector). Poxvirus vectors capable of expressing a polynucleotide of the invention can be obtained by homologous recombination, as described in Kieny *et al.* (Nature 312:163, 1984) so that the polynucleotide of the invention is inserted in the viral genome under appropriate conditions for expression in mammalian cells. Generally, the dose of viral vector vaccine, for therapeutic or prophylactic use, can be from about $1x10^4$ to about $1x10^{11}$, advantageously from about $1x10^7$ to about $1x10^{10}$, or, preferably, from about $1x10^7$ to about $1x10^9$ plaque-forming units per kilogram. Preferably, viral vectors are administered parenterally, for example, in 3 doses that are 4 weeks apart. Those skilled in the art will recognize that it is preferable to avoid adding a

chemical adjuvant to a composition containing a viral vector of the invention and thereby minimizing the immune response to the viral vector itself.

Non-toxicogenic *Vibrio cholerae* mutant strains that can be used in live oral vaccines are described by Mekalanos *et al.* (Nature 306:551, 1983) and in U.S. Patent No. 4,882,278 (strain in which a substantial amount of the coding sequence of each of the two *ctxA* alleles has been deleted so that no functional *cholerae* toxin is produced); WO 92/11354 (strain in which the *irg*A locus is inactivated by mutation; this mutation can be combined in a single strain with *ctx*A mutations); and WO 94/1533 (deletion mutant lacking functional *ctx*A and *attRS1* DNA sequences). These strains can be genetically engineered to express heterologous antigens, as described in WO 94/19482. An effective vaccine dose of a *V. cholerae* strain capable of expressing a polypeptide or polypeptide derivative encoded by a polynucleotide molecule of the invention can contain, *e.g.*, about $1x10^5$ to about $1x10^9$, preferably about $1x10^6$ to about $1x10^8$ viable bacteria in an appropriate volume for the selected route of administration. Preferred routes of administration include all mucosal routes, but, most preferably, these vectors are administered intranasally or orally.

Attenuated *Salmonella typhimurium* strains, genetically engineered for recombinant expression of heterologous antigens, and their use as oral vaccines, are described by Nakayama *et al.* (Bio/Technology 6:693, 1988) and in WO 92/11361. Preferred routes of administration for these vectors include all mucosal routes. Most preferably, the vectors are administered intranasally or orally.

Others bacterial strains useful as vaccine vectors are described by High *et al.* (EMBO 11:1991, 1992) and Sizemore *et al.* (Science 270:299, 1995; *Shigella flexneri*); Medaglini *et al.* (Proc. Natl. Acad. Sci. USA 92:6868,

1995; (*Streptococcus gordonii*); Flynn (Cell. Mol. Biol. 40 (suppl. I):31, 1194), and in WO 88/6626, WO 90/0594, WO 91/13157, WO 92/1796, and WO 92/21376 (Bacille Calmette Guerin). In bacterial vectors, a polynucleotide of the invention can be inserted into the bacterial genome or it can remain in a free state, for example, carried on a plasmid.

An adjuvant can also be added to a composition containing a bacterial vector vaccine. A number of adjuvants that can be used are known to those skilled in the art. For example, preferred adjuvants can be selected from the list provided below.

According to a fourth aspect of the invention, there is also provided (i) a composition of matter containing a polynucleotide of the invention, together with a diluent or carrier; (ii) a pharmaceutical composition containing a therapeutically or prophylactically effective amount of a polynucleotide of the invention; (iii) a method for inducing an immune response against *Helicobacter*, in a mammal, by administering to the mammal an immunogenically effective amount of a polynucleotide of the invention to elicit an immune response, *e.g.*, a protective immune response to *Helicobacter*; and (iv) a method for preventing and/or treating a *Helicobacter* (*e.g., H. pylori, H. felis, H. mustelae*, or *H. heilmanii*) infection, by administering a prophylactic or therapeutic amount of a polynucleotide of the invention to an individual in need of such treatment. Additionally, the fourth aspect of the invention encompasses the use of a polynucleotide of the invention in the preparation of a medicament for preventing and/or treating *Helicobacter* infection. The fourth aspect of the invention preferably includes the use of a polynucleotide molecule placed under conditions for expression in a mammalian cell, *e.g.*, in a plasmid that is unable to replicate in mammalian cells and to substantially integrate into a mammalian genome.

Polynucleotides (for example, DNA or RNA molecules) of the invention can also be administered as such to a mammal as a vaccine. When a DNA molecule of the invention is used, it can be in the form of a plasmid that is unable to replicate in a mammalian cell and unable to integrate into the mammalian genome. Typically, a DNA molecule is placed under the control of a promoter suitable for expression in a mammalian cell. The promoter can function ubiquitously or tissue-specifically. Examples of non-tissue specific promoters include the early Cytomegalovirus (CMV) promoter (U.S. Patent No. 4,168,062) and the Rous Sarcoma Virus promoter (Norton *et al.*, Molec. Cell Biol. 5:281, 1985). The desmin promoter (Li *et al.*, Gene 78:243, 1989; Li *et al.*, J. Biol. Chem. 266:6562, 1991; Li *et al.*, J. Biol. Chem. 268:10403, 1993) is tissue-specific and drives expression in muscle cells. More generally, useful promoters and vectors are described, *e.g.*, in WO 94/21797 and by Hartikka *et al.* (Human Gene Therapy 7:1205, 1996).

For DNA/RNA vaccination, the polynucleotide of the invention can encode a precursor or a mature form of a polypeptide of the invention. When it encodes a precursor form, the precursor sequence can be homologous or heterologous. In the latter case, a eucaryotic leader sequence can be used, such as the leader sequence of the tissue-type plasminogen factor (tPA).

A composition of the invention can contain one or several polynucleotides of the invention. It can also contain at least one additional polynucleotide encoding another *Helicobacter* antigen, such as urease subunit A, B, or both, or a fragment, derivative, mutant, or analog thereof. A polynucleotide encoding a cytokine, such as interleukin-2 (IL-2) or interleukin-12 (IL-12), can also be added to the composition so that the immune response is enhanced. These additional polynucleotides are placed under appropriate control for expression. Advantageously, DNA molecules of the invention

and/or additional DNA molecules to be included in the same composition are carried in the same plasmid.

Standard methods can be used in the preparation of therapeutic polynucleotides of the invention. For example, a polynucleotide can be used in a naked form, free of any delivery vehicles, such as anionic liposomes, cationic lipids, microparticles, *e.g.*, gold microparticles, precipitating agents, *e.g.*, calcium phosphate, or any other transfection-facilitating agent. In this case, the polynucleotide can be simply diluted in a physiologically acceptable solution, such as sterile saline or sterile buffered saline, with or without a carrier. When present, the carrier preferably is isotonic, hypotonic, or weakly hypertonic, and has a relatively low ionic strength, such as provided by a sucrose solution, *e.g.*, a solution containing 20% sucrose.

Alternatively, a polynucleotide can be associated with agents that assist in cellular uptake. It can be, *e.g.*, (i) complemented with a chemical agent that modifies cellular permeability, such as bupivacaine (see, *e.g.*, WO 94/16737), (ii) encapsulated into liposomes, or (iii) associated with cationic lipids or silica, gold, or tungsten microparticles.

Anionic and neutral liposomes are well-known in the art (see, *e.g.*, *Liposomes: A Practical Approach*, RPC New Ed, IRL Press, 1990, for a detailed description of methods for making liposomes) and are useful for delivering a large range of products, including polynucleotides.

Cationic lipids can also be used for gene delivery. Such lipids include, for example, Lipofectin™, which is also known as DOTMA (N-[1-(2,3-dioleyloxy)propyl]-N,N,N-trimethylammonium chloride), DOTAP (1,2-bis(oleyloxy)-3-(trimethylammonio)propane), DDAB (dimethyldioctadecylammonium bromide), DOGS (dioctadecylamidologlycyl spermine), and cholesterol derivatives. A description of these cationic lipids

can be found in EP 187,702, WO 90/11092, U.S. Patent No. 5,283,185, WO 91/15501, WO 95/26356, and U.S. Patent No. 5,527,928. Cationic lipids for gene delivery are preferably used in association with a neutral lipid such as DOPE (dioleyl phosphatidylethanolamine; WO 90/11092). Other transfection-facilitating compounds can be added to a formulation containing cationic liposomes. A number of them are described in, *e.g.*, WO 93/18759, WO 93/19768, WO 94/25608, and WO 95/2397. They include, *e.g.*, spermine derivatives useful for facilitating the transport of DNA through the nuclear membrane (see, for example, WO 93/18759) and membrane-permeabilizing compounds such as GALA, Gramicidine S, and cationic bile salts (see, for example, WO 93/19768).

Gold or tungsten microparticles can also be used for gene delivery, as described in WO 91/359, WO 93/17706, and by Tang *et al.* (Nature 356:152, 1992). In this case, the microparticle-coated polynucleotides can be injected *via* intradermal or intraepidermal routes using a needleless injection device ("gene gun"), such as those described in U.S. Patent No. 4,945,050, U.S. Patent No. 5,015,580, and WO 94/24263.

The amount of DNA to be used in a vaccine recipient depends, *e.g.*, on the strength of the promoter used in the DNA construct, the immunogenicity of the expressed gene product, the condition of the mammal intended for administration (*e.g.*, the weight, age, and general health of the mammal), the mode of administration, and the type of formulation. In general, a therapeutically or prophylactically effective dose from about 1 μg to about 1 mg, preferably, from about 10 μg to about 800 μg, and, more preferably, from about 25 μg to about 250 μg, can be administered to human adults. The administration can be achieved in a single dose or repeated at intervals.

The route of administration can be any conventional route used in the vaccine field. As general guidance, a polynucleotide of the invention can be administered *via* a mucosal surface, *e.g.*, an ocular, intranasal, pulmonary, oral, intestinal, rectal, vaginal, or urinary tract surface, or *via* a parenteral route, *e.g.*, by an intravenous, subcutaneous, intraperitoneal, intradermal, intraepidermal, or intramuscular route. The choice of administration route will depend on, *e.g.*, the formulation that is selected. A polynucleotide formulated in association with bupivacaine is advantageously administered into muscle. When a neutral or anionic liposome or a cationic lipid, such as DOTMA, is used, the formulation can be advantageously injected *via* intravenous, intranasal (for example, by aerosolization), intramuscular, intradermal, and subcutaneous routes. A polynucleotide in a naked form can advantageously be administered *via* the intramuscular, intradermal, or subcutaneous routes. Although not absolutely required, such a composition can also contain an adjuvant. A systemic adjuvant that does not require concomitant administration in order to exhibit an adjuvant effect is preferable.

The sequence information provided in the present application enables the design of specific nucleotide probes and primers that can be used in diagnostic methods. Accordingly, in a fifth aspect of the invention, there is provided a nucleotide probe or primer having a sequence found in, or derived by degeneracy of the genetic code from, a sequence shown in any of SEQ ID NOs:1-21 (odd numbers), 65, and 67, or a complementary sequence thereof.

The term "probe" as used in the present application refers to DNA (preferably single stranded) or RNA molecules (or modifications or combinations thereof) that hybridize under the stringent conditions, as defined above, to polynucleotide molecules having sequences homologous to those shown in any of SEQ ID NOs:1-21 (odd numbers), 65, and 67, or to a

complementary or anti-sense sequence of any of SEQ ID NOs:1-21 (odd numbers), 65, and 67. Generally, probes are significantly shorter than the full-length sequences shown in any of SEQ ID NOs:1-21 (odd numbers), 65, and 67. For example, they can contain from about 5 to about 100, preferably from about 10 to about 80 nucleotides. In particular, probes have sequences that are at least 75%, preferably at least 85%, more preferably 95% homologous to a portion of a sequence as shown in any of SEQ ID NOs:1-21 (odd numbers), 65, and 67, or a sequence complementary to such sequences.

Probes can contain modified bases, such as inosine, methyl-5-deoxycytidine, deoxyuridine, dimethylamino-5-deoxyuridine, or diamino-2, 6-purine. Sugar or phosphate residues can also be modified or substituted. For example, a deoxyribose residue can be replaced by a polyamide (Nielsen *et al.*, Science 254:1497, 1991) and phosphate residues can be replaced by ester groups such as diphosphate, alkyl, arylphosphonate, and phosphorothioate esters. In addition, the 2'-hydroxyl group on ribonucleotides can be modified by addition of, *e.g.*, alkyl groups.

Probes of the invention can be used in diagnostic tests, or as capture or detection probes. Such capture probes can be immobilized on solid supports, directly or indirectly, by covalent means or by passive adsorption. A detection probe can be labeled by a detectable label, for example a label selected from radioactive isotopes; enzymes, such as peroxidase and alkaline phosphatase; enzymes that are able to hydrolyze a chromogenic, fluorogenic, or luminescent substrate; compounds that are chromogenic, fluorogenic, or luminescent; nucleotide base analogs; and biotin.

Probes of the invention can be used in any conventional hybridization method, such as in dot blot methods (Maniatis *et al.*, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold

Spring Harbor, New York, 1982), Southern blot methods (Southern, J. Mol. Biol. 98:503, 1975), northern blot methods (identical to Southern blot to the exception that RNA is used as a target), or a sandwich method (Dunn *et al.*, Cell 12:23, 1977). As is known in the art, the latter technique involves the use of a specific capture probe and a specific detection probe that have nucleotide sequences that are at least partially different from each other.

Primers used in the invention usually contain about 10 to 40 nucleotides and are used to initiate enzymatic polymerization of DNA in an amplification process (*e.g.*, PCR), an elongation process, or a reverse transcription method. In a diagnostic method involving PCR, the primers can be labeled.

Thus, the invention also encompasses (i) a reagent containing a probe of the invention for detecting and/or identifying the presence of *Helicobacter* in a biological material; (ii) a method for detecting and/or identifying the presence of *Helicobacter* in a biological material, in which (a) a sample is recovered or derived from the biological material, (b) DNA or RNA is extracted from the material and denatured, and (c) the sample is exposed to a probe of the invention, for example, a capture probe, a detection probe, or both, under stringent hybridization conditions, so that hybridization is detected; and (iii) a method for detecting and/or identifying the presence of *Helicobacter* in a biological material, in which (a) a sample is recovered or derived from the biological material, (b) DNA is extracted therefrom, (c) the extracted DNA is contacted with at least one, or, preferably two, primers of the invention, and amplified by the polymerase chain reaction, and (d) an amplified DNA molecule is produced.

As mentioned above, polypeptides that can be produced by expression of the polynucleotides of the invention can be used as vaccine

antigens. Accordingly, a sixth aspect of the invention features a substantially purified polypeptide or polypeptide derivative having an amino acid sequence encoded by a polynucleotide of the invention.

A "substantially purified polypeptide" is defined as a polypeptide that is separated from the environment in which it naturally occurs and/or a polypeptide that is free of most of the other polypeptides that are present in the environment in which it was synthesized. The polypeptides of the invention can be purified from a natural source, such as a *Helicobacter* strain, or can be produced using recombinant methods.

Homologous polypeptides or polypeptide derivatives encoded by polynucleotides of the invention can be screened for specific antigenicity by testing cross-reactivity with an antiserum raised against a polypeptide having an amino acid sequence as shown in any of SEQ ID NOs:2-22 (even numbers), 66, and 68. Briefly, a monospecific hyperimmune antiserum can be raised against a purified reference polypeptide as such or as a fusion polypeptide, for example, an expression product of MBP, GST, or His-tag systems, or a synthetic peptide predicted to be antigenic. The homologous polypeptide or derivative that is screened for specific antigenicity can be produced as such or as a fusion polypeptide. In the latter case, and if the antiserum is also raised against a fusion polypeptide, two different fusion systems are employed. Specific antigenicity can be determined using a number of methods, including Western blot (Towbin *et al.*, Proc. Natl. Acad. Sci. USA 76:4350, 1979), dot blot, and ELISA methods, as described below.

In a Western blot assay, the product to be screened, either as a purified preparation or a total *E. coli* extract, is fractionated by SDS-PAGE, as described, for example, by Laemmli (Nature 227:680, 1970). After being transferred to a filter, such as a nitrocellulose membrane, the material is

incubated with the monospecific hyperimmune antiserum, which is diluted in a range of dilutions from about 1:50 to about 1:5000, preferably from about 1:100 to about 1:500. Specific antigenicity is shown once a band corresponding to the product exhibits reactivity at any of the dilutions in the range.

In an ELISA assay, the product to be screened can be used as the coating antigen. A purified preparation is preferred, but a whole cell extract can also be used. Briefly, about 100 µl of a preparation of about 10 µg protein/ml is distributed into wells of a 96-well ELISA plate. The plate is incubated for about 2 hours at 37°C, then overnight at 4°C. The plate is washed with phosphate buffered saline (PBS) containing 0.05% Tween 20 (PBS/Tween buffer) and the wells are saturated with 250 µl PBS containing 1% bovine serum albumin (BSA), to prevent non-specific antibody binding. After 1 hour of incubation at 37°C, the plate is washed with PBS/Tween buffer. The antiserum is serially diluted in PBS/Tween buffer containing 0.5% BSA, and 100 µl dilutions are added to each well. The plate is incubated for 90 minutes at 37°C, washed, and evaluated using standard methods. For example, a goat anti-rabbit peroxidase conjugate can be added to the wells when the specific antibodies used were raised in rabbits. Incubation is carried out for about 90 minutes at 37°C and the plate is washed. The reaction is developed with the appropriate substrate and the reaction is measured by colorimetry (absorbance measured spectrophotometrically). Under these experimental conditions, a positive reaction is shown once an O.D. value of 1.0 is detected with a dilution of at least about 1:50, preferably of at least about 1:500.

In a dot blot assay, a purified product is preferred, although a whole cell extract can be used. Briefly, a solution of the product at a concentration of

about 100 μg/ml is serially diluted two-fold with 50 mM Tris-HCl (pH 7.5). One hundred μl of each dilution is applied to a filter, such as a 0.45 μm nitrocellulose membrane, set in a 96-well dot blot apparatus (Biorad). The buffer is removed by applying vacuum to the system. Wells are washed by addition of 50 mM Tris-HCl (pH 7.5) and the membrane is air-dried. The membrane is saturated in blocking buffer (50 mM Tris-HCl (pH 7.5), 0.15 M NaCl, 10 g/L skim milk) and incubated with an antiserum diluted from about 1:50 to about 1:5000, preferably about 1:500. The reaction is detected using standard methods. For example, a goat anti-rabbit peroxidase conjugate can be added to the wells when rabbit antibodies are used. Incubation is carried out for about 90 minutes at 37°C and the blot is washed. The reaction is developed with the appropriate substrate and stopped. The reaction is then measured visually by the appearance of a colored spot, *e.g.*, by colorimetry. Under these experimental conditions, a positive reaction is associated with detection of a colored spot for reactions carried out with a dilution of at least about 1:50, preferably, of at least about 1:500. Therapeutic or prophylactic efficacy of a polypeptide or polypeptide derivative of the invention can be evaluated as is described below.

According to a seventh aspect of the invention, there is provided (i) a composition of matter containing a polypeptide of the invention together with a diluent or carrier; (ii) a pharmaceutical composition containing a therapeutically or prophylactically effective amount of a polypeptide of the invention; (iii) a method for inducing an immune response against *Helicobacter* in a mammal by administering to the mammal an immunogenically effective amount of a polypeptide of the invention to elicit an immune response, *e.g.*, a protective immune response to *Helicobacter*; and (iv) a method for preventing and/or treating a *Helicobacter* (*e.g., H. pylori, H. felis, H. mustelae*, or *H.*

heilmanii) infection, by administering a prophylactic or therapeutic amount of a polypeptide of the invention to an individual in need of such treatment. Additionally, this aspect of the invention includes the use of a polypeptide of the invention in the preparation of a medicament for preventing and/or treating *Helicobacter* infection.

The immunogenic compositions of the invention can be administered by any conventional route in use in the vaccine field, for example, to a mucosal (*e.g.*, ocular, intranasal, pulmonary, oral, gastric, intestinal, rectal, vaginal, or urinary tract) surface or *via* a parenteral (*e.g.*, subcutaneous, intradermal, intramuscular, intravenous, or intraperitoneal) route. The choice of the administration route depends upon a number of parameters, such as the adjuvant used. For example, if a mucosal adjuvant is used, the intranasal or oral route will be preferred, and if a lipid formulation or an aluminum compound is used, a parenteral route will be preferred. In the latter case, the subcutaneous or intramuscular route is most preferred. The choice of administration route can also depend upon the nature of the vaccine agent. For example, a polypeptide of the invention fused to CTB or to LTB will be best administered to a mucosal surface.

A composition of the invention can contain one or several polypeptides or derivatives of the invention. It can also contain at least one additional *Helicobacter* antigen, such as the urease apoenzyme, or a subunit, fragment, homolog, mutant, or derivative thereof.

For use in a composition of the invention, a polypeptide or polypeptide derivative can be formulated into or with liposomes, such as neutral or anionic liposomes, microspheres, ISCOMS, or virus-like particles (VLPs), to facilitate delivery and/or enhance the immune response. These compounds are readily available to those skilled in the art; for example, see

Liposomes: A Practical Approach (*supra*). Adjuvants other than liposomes can also be used in the invention and are well known in the art (see, for example, the list provided below).

Administration can be achieved in a single dose or repeated as necessary at intervals that can be determined by one skilled in the art. For example, a priming dose can be followed by three booster doses at weekly or monthly intervals. An appropriate dose depends on various parameters, including the nature of the recipient (*e.g.*, whether the recipient is an adult or an infant), the particular vaccine antigen, the route and frequency of administration, the presence/absence or type of adjuvant, and the desired effect (*e.g.*, protection and/or treatment), and can be readily determined by one skilled in the art. In general, a vaccine antigen of the invention can be administered mucosally in an amount ranging from about 10 μg to about 500 mg, preferably from about 1 mg to about 200 mg. For a parenteral route of administration, the dose usually should not exceed about 1 mg, and is, preferably, about 100 μg.

When used as components of a vaccine, the polynucleotides and polypeptides of the invention can be used sequentially as part of a multi-step immunization process. For example, a mammal can be initially primed with a vaccine vector of the invention, such as a pox virus, *e.g.*, *via* a parenteral route, and then boosted twice with a polypeptide encoded by the vaccine vector, *e.g.*, *via* the mucosal route. In another example, liposomes associated with a polypeptide or polypeptide derivative of the invention can be used for priming, with boosting being carried out mucosally using a soluble polypeptide or polypeptide derivative of the invention, in combination with a mucosal adjuvant (*e.g.*, LT).

Polypeptides and polypeptide derivatives of the invention can also be used as diagnostic reagents for detecting the presence of anti-*Helicobacter*

antibodies, *e.g.*, in blood samples. Such polypeptides can be about 5 to about 80, preferably, about 10 to about 50 amino acids in length and can be labeled or unlabeled, depending upon the diagnostic method. Diagnostic methods involving such a reagent are described below.

Upon expression of a polynucleotide molecule of the invention, a polypeptide or polypeptide derivative is produced and can be purified using known methods. For example, the polypeptide or polypeptide derivative can be produced as a fusion protein containing a fused tail that facilitates purification. The fusion product can be used to immunize a small mammal, *e.g.*, a mouse or a rabbit, in order to raise monospecific antibodies against the polypeptide or polypeptide derivative. The eighth aspect of the invention thus provides a monospecific antibody that binds to a polypeptide or polypeptide derivative of the invention.

By "monospecific antibody" is meant an antibody that is capable of reacting with a unique, naturally-occurring *Helicobacter* polypeptide. An antibody of the invention can be polyclonal or monoclonal. Monospecific antibodies can be recombinant, *e.g.*, chimeric (*e.g.*, consisting of a variable region of murine origin and a human constant region), humanized (*e.g.*, a human immunoglobulin constant region and a variable region of animal, *e.g.*, murine, origin), and/or single chain. Both polyclonal and monospecific antibodies can also be in the form of immunoglobulin fragments, *e.g.*, F(ab)'2 or Fab fragments. The antibodies of the invention can be of any isotype, *e.g.*, IgG or IgA, and polyclonal antibodies can be of a single isotype or can contain a mixture of isotypes.

The antibodies of the invention, which can be raised to a polypeptide or polypeptide derivative of the invention, can be produced and identified using standard immunological assays, *e.g.*, Western blot assays, dot blot assays, or

ELISA (see, *e.g.*, Coligan *et al.*, *Current Protocols in Immunology*, John Wiley & Sons, Inc., New York, NY, 1994). The antibodies can be used in diagnostic methods to detect the presence of *Helicobacter* antigens in a sample, such as a biological sample. The antibodies can also be used in affinity chromatography methods for purifying a polypeptide or polypeptide derivative of the invention. As is discussed further below, the antibodies can also be used in prophylactic and therapeutic passive immunization methods.

Accordingly, a ninth aspect of the invention provides (i) a reagent for detecting the presence of *Helicobacter* in a biological sample that contains an antibody, polypeptide, or polypeptide derivative of the invention; and (ii) a diagnostic method for detecting the presence of *Helicobacter* in a biological sample, by contacting the biological sample with an antibody, a polypeptide, or a polypeptide derivative of the invention, so that an immune complex is formed, and detecting the complex as an indication of the presence of *Helicobacter* in the sample or the organism from which the sample was derived. The immune complex is formed between a component of the sample and the antibody, polypeptide, or polypeptide derivative, and that any unbound material can be removed prior to detecting the complex. A polypeptide reagent can be used for detecting the presence of anti-*Helicobacter* antibodies in a sample, *e.g.*, a blood sample, while an antibody of the invention can be used for screening a sample, such as a gastric extract or biopsy sample, for the presence of *Helicobacter* polypeptides.

For use in diagnostic methods, the reagent (*e.g.*, the antibody, polypeptide, or polypeptide derivative of the invention) can be in a free state or can be immobilized on a solid support, such as, for example, on the interior surface of a tube or on the surface, or within pores, of a bead. Immobilization can be achieved using direct or indirect means. Direct means include passive

adsorption (*i.e.*, non-covalent binding) or covalent binding between the support and the reagent. By "indirect means" is meant that an anti-reagent compound that interacts with the reagent is first attached to the solid support. For example, if a polypeptide reagent is used, an antibody that binds to it can serve as an anti-reagent, provided that it binds to an epitope that is not involved in recognition of antibodies in biological samples. Indirect means can also employ a ligand-receptor system, for example, a molecule, such as a vitamin, can be grafted onto the polypeptide reagent and the corresponding receptor can be immobilized on the solid phase. This concept is illustrated by the well known biotin-streptavidin system. Alternatively, indirect means can be used, *e.g.*, by adding to the reagent a peptide tail, chemically or by genetic engineering, and immobilizing the grafted or fused product by passive adsorption or covalent linkage of the peptide tail.

According to a tenth aspect of the invention, there is provided a process for purifying, from a biological sample, a polypeptide or polypeptide derivative of the invention, which involves carrying out antibody-based affinity chromatography with the biological sample, wherein the antibody is a monospecific antibody of the invention.

For use in a purification process of the invention, the antibody can be polyclonal or monospecific, and preferably is of the IgG type. Purified IgGs can be prepared from an antiserum using standard methods (see, *e.g.*, Coligan *et al.*, *supra*). Conventional chromatography supports, as well as standard methods for grafting antibodies, are described, for example, by Harlow *et al.* (*Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1988).

Briefly, a biological sample, such as an *H. pylori* extract, preferably in a buffer solution, is applied to a chromatography material, which is,

preferably, equilibrated with the buffer used to dilute the biological sample, so that the polypeptide or polypeptide derivative of the invention (*i.e.*, the antigen) is allowed to adsorb onto the material. The chromatography material, such as a gel or a resin coupled to an antibody of the invention, can be in batch form or in a column. The unbound components are washed off and the antigen is eluted with an appropriate elution buffer, such as a glycine buffer, a buffer containing a chaotropic agent, *e.g.*, guanidine HCl, or a buffer having high salt concentration (*e.g.*, 3 M $MgCl_2$). Eluted fractions are recovered and the presence of the antigen is detected, *e.g.*, by measuring the absorbance at 280 nm.

An antibody of the invention can be screened for therapeutic efficacy as follows. According to an eleventh aspect of the invention, there is provided (i) a composition of matter containing a monospecific antibody of the invention, together with a diluent or carrier; (ii) a pharmaceutical composition containing a therapeutically or prophylactically effective amount of a monospecific antibody of the invention; and (iii) a method for treating or preventing *Helicobacter* (*e.g., H. pylori, H. felis, H. mustelae*, or *H. heilmanii*) infection, by administering a therapeutic or prophylactic amount of a monospecific antibody of the invention to an individual in need of such treatment. In addition, the eleventh aspect of the invention includes the use of a monospecific antibody of the invention in the preparation of a medicament for treating or preventing *Helicobacter* infection.

The monospecific antibody can be polyclonal or monoclonal, and is, preferably, predominantly of the IgA isotype. In passive immunization methods, the antibody is administered to a mucosal surface of a mammal, *e.g.*, the gastric mucosa, *e.g.*, orally or intragastrically, optionally, in the presence of a bicarbonate buffer. Alternatively, systemic administration, not requiring a

bicarbonate buffer, can be carried out. A monospecific antibody of the invention can be administered as a single active agent or as a mixture with at least one additional monospecific antibody specific for a different *Helicobacter* polypeptide. The amount of antibody and the particular regimen used can be readily determined by one skilled in the art. For example, daily administration of about 100 to 1,000 mg of antibody over one week, or three doses per day of about 100 to 1,000 mg of antibody over two or three days, can be effective regimens for most purposes.

Therapeutic or prophylactic efficacy can be evaluated using standard methods in the art, *e.g.*, by measuring induction of a mucosal immune response or induction of protective and/or therapeutic immunity, using, *e.g.*, the *H. felis* mouse model and the procedures described by Lee *et al.* (Eur. J. Gastroenterology & Hepatology 7:303, 1995) or Lee *et al.* (J. Infect. Dis. 172:161, 1995). Those skilled in the art will recognize that the *H. felis* strain of the model can be replaced with another *Helicobacter* strain. For example, the efficacy of polynucleotide molecules and polypeptides from *H. pylori* is, preferably, evaluated in a mouse model using an *H. pylori* strain. Protection can be determined by comparing the degree of *Helicobacter* infection in the gastric tissue assessed by, for example, urease activity, bacterial counts, or gastritis, to that of a control group. Protection is shown when infection is reduced by comparison to the control group. Such an evaluation can be made for polynucleotides, vaccine vectors, polypeptides, and polypeptide derivatives, as well as for antibodies of the invention.

For example, various doses of an antibody of the invention can be administered to the gastric mucosa of mice previously challenged with an *H. pylori* strain, as described, *e.g.*, by Lee *et al.* (*supra*). Then, after an appropriate period of time, the bacterial load of the mucosa can be estimated by

assessing urease activity, as compared to a control. Reduced urease activity indicates that the antibody is therapeutically effective.

Adjuvants that can be used in any of the vaccine compositions described above are described as follows. Adjuvants for parenteral administration include, for example, aluminum compounds, such as aluminum hydroxide, aluminum phosphate, and aluminum hydroxy phosphate. The antigen can be precipitated with, or adsorbed onto, the aluminum compound using standard methods. Other adjuvants, such as RIBI (ImmunoChem, Hamilton, MT), can also be used in parenteral administration.

Adjuvants that can be used for mucosal administration include, for example, bacterial toxins, *e.g.*, the cholera toxin (CT), the *E. coli* heat-labile toxin (LT), the *Clostridium difficile* toxin A, the *pertussis* toxin (PT), and combinations, subunits, toxoids, or mutants thereof. For example, a purified preparation of native cholera toxin subunit B (CTB) can be used. Fragments, homologs, derivatives, and fusions to any of these toxins can also be used, provided that they retain adjuvant activity. Preferably, a mutant having reduced toxicity is used. Suitable mutants are described, *e.g.*, in WO 95/17211 (Arg-7-Lys CT mutant), WO 96/6627 (Arg-192-Gly LT mutant), and WO 95/34323 (Arg-9-Lys and Glu-129-Gly PT mutant). Additional LT mutants that can be used in the methods and compositions of the invention include, *e.g.*, Ser-63-Lys, Ala-69-Gly, Glu-110-Asp, and Glu-112-Asp mutants. Other adjuvants, such as the bacterial monophosphoryl lipid A (MPLA) of, *e.g., E. coli, Salmonella minnesota, Salmonella typhimurium,* or *Shigella flexneri*; saponins, and polylactide glycolide (PLGA) microspheres, can also be used in mucosal administration. Adjuvants useful for both mucosal and parenteral administrations, such as polyphosphazene (WO 95/2415), can also be used.

Any pharmaceutical composition of the invention, containing a polynucleotide, polypeptide, polypeptide derivative, or antibody of the invention, can be manufactured using standard methods. It can be formulated with a pharmaceutically acceptable diluent or carrier, *e.g.*, water or a saline solution, such as PBS, optionally, including a bicarbonate salt, such as sodium bicarbonate, *e.g.*, 0.1 to 0.5 M. Bicarbonate can advantageously be added to compositions intended for oral or intragastric administration. In general, a diluent or carrier can be selected on the basis of the mode and route of administration, and standard pharmaceutical practice. Suitable pharmaceutical carriers and diluents, as well as pharmaceutical necessities for their use in pharmaceutical formulations, are described in *Remington's Pharmaceutical Sciences*, a standard reference text in this field and in the USP/NF.

The invention also includes methods in which gastroduodenal infections, such as *Helicobacter* infection, are treated by oral administration of a *Helicobacter* polypeptide of the invention and a mucosal adjuvant, in combination with an antibiotic, an antisecretory agent, a bismuth salt, an antacid, sucralfate, or a combination thereof. Examples of such compounds that can be administered with the vaccine antigen and an adjuvant are antibiotics, including, *e.g.*, macrolides, tetracyclines, β-lactams, aminoglycosides, quinolones, penicillins, and derivatives thereof (specific examples of antibiotics that can be used in the invention include, *e.g.*, amoxicillin, clarithromycin, tetracycline, metronidizole, erythromycin, cefuroxime, and erythromycin); antisecretory agents, including, *e.g.*, H_2-receptor antagonists (*e.g.*, cimetidine, ranitidine, famotidine, nizatidine, and roxatidine), proton pump inhibitors (*e.g.*, omeprazole, lansoprazole, and pantoprazole), prostaglandin analogs (*e.g.*, misoprostil and enprostil), and anticholinergic agents (*e.g.*, pirenzepine, telenzepine, carbenoxolone, and

proglumide); and bismuth salts, including colloidal bismuth subcitrate, tripotassium dicitrate bismuthate, bismuth subsalicylate, bicitropeptide, and pepto-bismol (see, *e.g.*, Goodwin *et al.*, *Helicobacter pylori, Biology and Clinical Practice*, CRC Press, Boca Raton, FL, pp 366-395, 1993; Physicians' Desk Reference, 49th edn., Medical Economics Data Production Company, Montvale, New Jersey, 1995). In addition, compounds containing more than one of the above-listed components coupled together, *e.g.*, ranitidine coupled to bismuth subcitrate, can be used. The invention also includes compositions for carrying out these methods, *i.e.*, compositions containing a *Helicobacter* antigen (or antigens) of the invention, an adjuvant, and one or more of the above-listed compounds, in a pharmaceutically acceptable carrier or diluent.

Amounts of the above-listed compounds used in the methods and compositions of the invention can readily be determined by one skilled in the art. In addition, one skilled in the art can readily design treatment/immunization schedules. For example, the non-vaccine components can be administered on days 1-14, and the vaccine antigen + adjuvant can be administered on days 7, 14, 21, and 28.

Methods and pharmaceutical compositions of the invention can be used to treat or to prevent *Helicobacter* infections and, accordingly, gastroduodenal diseases associated with these infections, including acute, chronic, and atrophic gastritis, and peptic ulcer diseases, *e.g.*, gastric and duodenal ulcers.

A 76 kDa protein band containing GHPO 386, GHPO 789, and GHPO 1516 (hereinafter the "purified 76 kDa proteins"), GHPO 1360, and GHPO 750 were purified from *Helicobacter pylori* strain ATCC number 43579 (American Type Culture Collection, Rockville, Maryland) by immunoaffinity-

based chromatography using the methods described below in Example 1, and were shown to be effective vaccine antigens as follows.

Groups of 10 mice each were orally immunized with 1, 5, or 25 μg of the purified 76 kDa proteins, purified GHPO 1360, or purified GHPO 750 in combination with 5 μg of the heat-labile enterotoxin (LT) of *E. coli*. Twenty five μg of recombinant urease, in combination with 5 μg LT, was used as a positive control, and 5 μg of LT in PBS was used as a negative control. The immunizations were carried out four times each, on days 0, 7, 14, and 21 of the experiment. On day 33, blood samples were collected from the mice and, on day 34, saliva samples were collected. On day 35, all of the mice were challenged by intragastric administration of 1×10^7 streptomycin-resistant, mouse-adapted *H. pylori*. On day 49, additional saliva samples were collected and, about two weeks after challenge, on days 52-53, the mice were sacrificed. Stomachs were removed from the mice and were analyzed for *Helicobacter* infection by measuring urease activity in the intact stomach tissue and by a quantitative culture study (Table 1).

Briefly, these studies showed that the gastric urease activities in samples from mice immunized with all three amounts of the purified 76 kDa proteins (*i.e.*, 1, 5, and 25 μg), in combination with LT, were generally lower than the gastric urease activities of samples from mice immunized with LT alone or mice that were not treated prior to challenge. Levels of gastric urease activity generally decreased with increasing amounts of the protein administered, with the gastric urease activity levels for the 25 μg doses generally approaching those of mice immunized with 25 μg of recombinant urease and LT.

The quantitative culture analyses showed that the levels of *Helicobacter* detected in the stomachs of mice immunized with the purified 76

kDa proteins, purified GHPO 1360, or purified GHPO 750, which generally decreased with increasing dosages, were less than the levels detected in the stomachs of control mice that were immunized with LT alone or untreated before *Helicobacter* challenge (Tables 1 and 2). The percentages of mice protected by immunization with the purified 76 kDa proteins, purified GHPO 1360, or purified GHPO 750 met or approached the percentages of mice protected by treatment with urease (Tables 1 and 2). These results show that the purified 76 kDa proteins, GHPO 1360, and GHPO 750 are effective vaccine antigens for use in preventing *Helicobacter* infection.

Table 1

Prophylactic Immunization with PMsv Antigens as Oral Dose Response Against *H. pylori* Challenge

Treatment	BALB/c mice # mice infected Antrum (based on quantitative A_{550}, 0.148 O.D. cutoff)	Fisher's exact test infection status (based on quantitative A_{550} ratios, treatment group v. LT only (group 11)) p-value	CFU/ml (1/4 antrum) Mean ± SD	Wilcoxon rank sums test CFU treatment group v. LT only control (group 11) p-value
1 µg 50 kDa + LT	60% (6/10)	0.3034	30825 ± 23210	0.1736
5 µg 50 kDa + LT	40% (4/10)	0.0573	18910 ± 16341	0.0588
25 µg 50 kDa + LT	30% (3/10)	0.0198	22710 ± 32397	0.0821
1 µg 32 kDa + LT	50% (5/10)	0.1409	44225 ± 87824	0.0756
5 µg 50 kDa + LT	10% (1/10)	0.0011	11811 ± 11579	0.0191
25 µg 50 kDa + LT	0 (0/9)	0.0001	1608 ± 23917	0.0114
25 µg rUre + LT	0 (0/9)	0.0001	8208 ± 8021	0.0179
LT	90% (9/10) 90% (9/10)	- not determined	107340 ± 127949 46173 ± 42325	- 0.2568

Table 2

Prophylactic Immunization with PMsv Antigens as Oral Dose Response Against *H. pylori* Challenge				
Treatment	BALB/c mice # mice infected Antrum (based on quantitative A_{550}, 0.148 O.D. cutoff)	Fisher's exact test infection status (based on quantitative A_{550} ratios, treatment group v. LT only (group 11)) p-value	CFU/ml (1/4 antrum) Mean ± SD	Wilcoxon rank sums test CFU treatment group v. LT only control (group 11) p-value
1 µg 76 kDa + LT	56% (5/9)	0.1409	39922 ± 34708	0.2203
5 µg 76 kDa + LT	80% (4/5)	1	8802 ± 7788	0.0864
25 µg 76 kDa + LT	33% (3/9)	0.0198	9712 ± 12183	0.0178
25 µg rUre + LT	0 (0/9)	0.0001	8208 ± 8021	0.0179
LT	90% (9/10) 90% (9/10)	- not determined	107340 ± 127949 46173 ± 42325	- 0.2568

The invention is further illustrated by the following examples. Example 1 describes purification of GHPO 1516 (76 kDa), GHPO 1360 (32 kDa), and GHPO 750 (50 kDa) from *Helicobacter* cultures. Example 2 describes identification of genes, *e.g.*, genes encoding 76 kDa proteins, such as GHPO 386, GHPO 789, GHPO 1516, GHPO 1197, GHPO 1180, GHPO 896, GHPO 711, GHPO 190, GHPO 185, GHPO 1417, and GHPO 1414, a 32 kDa protein (GHPO 1360), and a 50 kDa protein (GHPO 750) in the *Helicobacter* genome, as well as identification of signal sequences, and primer design for amplification of genes lacking signal sequences. Example 3 describes cloning of DNA encoding GHPO 386, GHPO 789, GHPO 1516, GHPO 896, GHPO 1360, and GHPO 750 into a vector that provides a histidine tag, and production and purification of the resulting his-tagged fusion proteins. Example 4 describes methods for cloning DNA encoding the polypeptides of the invention so that they can be produced without His-tags, Example 5 describes methods

for purifying recombinant polypeptides of the invention, and Example 6 describes use of the GHPO 1360 polypeptide as a serodiagnostic tool for *H. pylori* infection

EXAMPLE 1: Purification and partial sequence analysis of GHPO 1516 (76 kDa), GHPO 1360 (32 kDa), and GHPO 750 (50 kDa) protein from *Helicobacter pylori*

1.A. Culture and initial purification steps

Frozen seeds from *H. pylori* strain ATCC 43579 are used to seed a 75 cm^2 flask containing a biphasic medium (a solid phase made of Colombia gelose containing 6% fresh sheep blood and a liquid phase made of triptcase soja containing 20% fetal calf serum). After 24 hours of culturing under microaerophilic conditions, the liquid phase is used for seeding several 75 cm^2 flasks containing biphasic medium lacking sheep blood. After 24 hours of culture, the liquid phase is used to seed a 2 L biofermentor in triptcase soja liquid phase containing 10 g/L beta-cyclodextrine. At OD 1.5-1.8, this culture is diluted in a 10 L biofermentor containing the liquid medium. After 24 hours, the bacteria are spun in a centrifuge at 4,000 x g for 30 minutes at 4°C. A 10 L culture contains about 20 to 30 g (wet weight) bacteria.

The pellet obtained using the method described above is washed with 500 ml PBS (7.650 g NaCl, 0.724 g disodium phosphate, and 0.210 g monopotassium phosphate for one liter (pH 7.2)) for a one liter culture. The bacteria are then spun in a centrifuge again under the same conditions.

The pellet (C1) is suspended in 1% N-octyl-D-glucopyranoside (NOG; 30 ml/L; Sigma). The bacterial suspension is incubated for 1 hour at

room temperature while stirring, spun in a centrifuge at 17,600 x g for 30 minutes at 4°C, and the pellet (C2) is recovered.

The supernatant (S2) is dialyzed against PBS overnight at 4°C while stirring. The precipitate is recovered by centrifugation at 2,600 x g for 30 minutes at 4°C. The supernatant (S2d) is discarded and the pellet (Cs2d) is recovered and stored at -20°C.

The pellet (C2) is resuspended in 20 mM Tris-HCl buffer (pH 7.5) and

100 μM Pefabloc (Buffer A), and is homogenized with an ultra-turrax (3821, Janke and Kungel). Lysozyme and EDTA are added at 0.1 mg/ml and 1 mM, respectively.

The homogenate is sonicated three times for 2 minutes each at 4°C, and then is spun in an ultracentrifuge at 210,000 x g for 30 minutes at 4°C. The supernatant (S3), which contains the cytoplasmic and periplasmic proteins, is eliminated, while the pellet is recovered, washed with buffer A, and spun in an ultracentrifuge at 210,000 x g for 30 minutes at 4°C. The supernatant (S4) is eliminated and the pellet (C4) is stored at -20°C. This pellet (C4) contains membrane proteins.

The pellet (C4) is washed in 50 mM $NaCO_3$ (pH 9.5) and 100 μM Pefabloc (buffer B). The suspension is spun in an ultracentrifuge at 210,000 x g for 30 minutes at 4°C. The supernatant (S5) is eliminated, and the pellet (C5) is then washed and spun in an ultracentrifuge as is described above. The supernatant (S6) is eliminated and the pellet (C6) is stored at -20°C.

1.B. Purification of the proteins of membrane fraction C4 by preparative SDS-PAGE

SDS-PAGE is carried out according to the method of Laemmli (*supra*), using a biphasic gel consisting of a 5% polyacrylamide concentrating gel and a 10% polyacrylamide separating gel. The membrane fraction C4 is resuspended in buffer A, diluted in an equal volume of 2x sample buffer, and heated for 5 minutes at 95°C. About 19 mg of protein is applied to the gel (16 x 12 cm; 5 mm thick). Pre-migration is carried out for 2 hours at 50 V, and is followed by migration overnight at 65 V. After Coomassie blue staining, five major bands are revealed that have apparent molecular weights of 87, 76, 54, 50, and 32 kDa. Bands at 50 and 32 kDa appear to be slightly contaminated with bands at 47 and 35 kDa, respectively.

A band corresponding to the purified 76 kDa proteins, 32 kDa protein (GHPO 1360), or 50 kDa protein (GHPO 750) is cut out from the gel and is pounded with an ultra-turrax in 10-20 ml extraction buffer (25 mM Tris-HCl (pH 8.8), 8 M urea, 10% SDS, 100 μM phenyl methyl sulfonyl fluoride (PMSF), and 10 μM Pefabloc (buffer C)).

Each homogenate is filtered through a Millipore AP20 filter under 7 bars at room temperature, washed with 5-10 ml buffer C, and then filtered again. Each filtrate is precipitated with three volumes of a 50/50 mixture of 75% methanol and 75% isopropanol, and then is spun in a centrifuge at 240,000 x g for 16 hours at 10°C.

Each pellet is resuspended in 2 ml of 10 mM $NaPO_4$ (pH 7.0) containing 1 M NaCl, 0.1% Sarkosyl, 100 μM PMSF, and 6 M urea (buffer D). The solubilized sample is dialyzed, in order, against 100 ml buffer D containing 4 M urea, 100 ml buffer D containing 2 M urea and 0.5% Sarkosyl, and twice against 100 ml buffer D that does not contain urea or Sarkosyl. The dialyses

are carried out for 1 hour each while stirring at room temperature. The last dialysate is incubated for 30 minutes in an ice bath, and then is spun in a centrifuge at low speed for 10 minutes at 4°C. The supernatant is recovered, filtered through a Millipore filter (0.45 μm), and stored at -20°C.

1.C. Purification of the 76 kDa, 32 kDa, or 50 kDa protein by immunoaffinity-based chromatography

1.C.1. Antiserum preparation

Specific polyclonal serum against the purified 76 kDa proteins, the 32 kDa protein (GHPO 1360), or the 50 kDa protein (GHPO 750), which are purified by preparative SDS-PAGE, is prepared by hyperimmunizing rabbits as follows. On day 0, a preparation containing 50 μg of the protein mixed with complete Freund's adjuvant is administered subcutaneously to the rabbits at multiple sites. The rabbits are boosted at days 21 and 42 with 25 μg of the protein in incomplete Freund's adjuvant, and are sacrificed at day 60. Complement is removed from the serum by heating for 30 minutes at 56°C. The hyperimmune serum is then sterilized by filtration through a Millipore membrane (0.22 μm).

1.C.2. IgG purification

The hyperimmune serum prepared as described above is applied to a Protein A Sepharose Fast Flow column (Pharmacia) that is equilibrated with 100 mM Tris-HCl (pH 8.0). The column is washed with 10 column volumes of 100 mM Tris-HCl (pH 8.0), and then with 10 column volumes of 10 mM Tris-HCl (pH 8.0). IgGs are eluted in 0.1 M glycine buffer (pH 3.0), and are collected as 5 ml fractions, to each of which 0.25 ml of Tris-HCl (pH 8.0) is added. The optical density of each fraction is measured at 280 nm, the IgG-containing fractions are pooled together and, if necessary, frozen at -70°C.

1.C.3. Preparation of the column

An appropriate amount of CNBr-activated Sepharose 4B gel (Pharmacia; reference: 17-0430-01) is suspended in 1 mM NaCl buffer (1 g dry gel provides for 3.5 ml hydrated gel; 5 to 10 mg IgGs can be retained per ml of hydrated gel). The gel is then washed using a buchner by adding small quantities of 1 mM HCl. The total volume of 1 mM HCl that is used amounts to 200 ml/g of gel.

Purified IgGs are dialyzed for 4 hours at room temperature against 50 volumes of 500 mM sodium phosphate buffer (pH 7.5). The IgGs are then diluted to 3 mg/ml with the same buffer. IgGs are incubated with the gel overnight at 5±3°C while stirring. The gel is packed in a chromatography column and is washed with 2 column volumes of 500 mM phosphate buffer (pH 7.5). The gel is then transferred to a tube and is incubated with 100 mM ethanolamine (pH 7.5), and then it is washed with 2 column volumes of PBS. The gel can be stored in PBS/merthiolate, 1/10,000.

1.C.4. Adsorption and elution

The 76 kDa protein is adsorbed and eluted as follows. The membrane fraction Cs2d is suspended in 50 mM Tris-HCl (pH 8.0), 2 mM EDTA, and then is filtered through a 0.45 μm membrane. The supernatant is applied to the column, which is equilibrated with 50 mM Tris-HCl (pH 8.0), 2 mM EDTA, at a flow rate of about 10 ml/hour. The column is washed with 20 column volumes of 50 mM Tris-HCl (pH 8.0), 2 mM EDTA, and then with 2 to 6 volumes 10 mM phosphate buffer (pH 6.8).

The antigen is eluted with 100 mM glycine buffer (pH 2.5). The eluate is collected in 3 ml fractions, to each of which is added 150 μl 1 M phosphate buffer (pH 8.0). The optical density of each fraction is measured at

280 nm, fractions containing the 76 kDa protein are pooled, and stored at -70°C.

Analysis by 10% SDS-PAGE reveals a single band at 76 kDa. N-terminal sequence was carried out on this purified 76 kDa preparation, and the sequence obtained is as follows: EDDGFYTSVGYQIGEAAQMV (SEQ ID NO:58).

The 32 kDa protein (GHPO 1360) or the 50 kDa protein (GHPO 750) is purified by immunoaffinity-based chromatography as follows. In order to separate the 32 or 50 kDa protein from the contaminating proteins (the 47 and 35 kDa proteins, respectively), membrane fraction C4 is solubilized in 50 mM $NaCO_3$ (pH 9.5) for 30 minutes at room temperature under stirring and the preparation is centrifuged for 30 minutes at 200,000 x g at 4°C. The 47 and 35 kDa proteins are insoluble in the $NaCO_3$ buffer and are eliminated in the pellet.

The supernatant is dialyzed against 50 mM Tris-HCL (pH 8.0), 2 mM EDTA, and then is filtered through a 0.45 μm membrane. The filtered supernatant is applied to the column, which is equilibrated with 50 mM Tris-HCL (pH 8.0), 2 mM EDTA, at a flow rate of about 10 ml/hour. The column is washed with 20 column volumes of 50 mM Tris-HCL (pH 8.0), 2 mM EDTA, and then with 2 to 6 volumes of 10 mM phosphate buffer (pH 6.8).

The antigen is eluted with 100 mM glycine buffer (pH 2.5). The eluate is collected in 3 ml fractions, to each of which is added 150 μl 1 M phosphate buffer (pH 8.0). The optical density of each fraction is measured at 280 nm, and fractions containing the 50 or 32 kDa protein are pooled and stored at -70°C.

Analysis of the purified protein by 10% SDS-PAGE reveals single bands at 50 and 32 kDa. N-terminal sequencing is carried out with the purified 50 kDa protein preparation. The sequence found is as follows: MKEKFNRTKPHVNIGTIGHVDH (SEQ ID NO:73). Similarly, N-terminal

and internal sequencing is carried out with the purified 32 kDa preparation. The sequences found are as follows: AHNANNATHNTKK (SEQ ID NO:74) and KPAHNA (SEQ ID NO:75) (N-terminal), and IDKQPKAKK (SEQ ID NO:76) and FWAKKQAE (SEQ ID NO:77) (internal).

1.D. Purification of the 76 kDa protein from membrane fraction Cs2d and purification of the 32 kDa and 50 kDa proteins from membrane fraction C4

The 76 kDa protein can also be purified as follows. A 40 ml Q-Sepharose column (diameter: 2.5 cm; height: 8 cm) is prepared according to the manufacturer's instructions (Pharmacia). The column is washed and equilibrated with buffer B, containing 50 mM $NaC0_3$ (pH 9.5), 100 μM Pefabloc, and 0.1% Zwittergent 3-14. The chromatography is monitored by measuring absorbance at 280 nm at the column exit.

One hundred and forty mg of protein from the membrane fraction Cs2d resuspended in buffer B are applied to the column. The column is washed with 0.1 M NaCl in buffer B, and then a 0.1-0.5 M NaCl gradient is applied to the column. The fraction eluted between 0.35 and 0.45 M NaCl is further purified on a 10 ml S-Sepharose column (diameter: 1.5 cm; height: 5 cm; up to 10 mg protein/ml of gel), which is prepared according to the manufacturer's instructions (Pharmacia). The fraction obtained is dialyzed against 50 mM acetate (pH 5.0) containing 100 μM Pefabloc and 0.1% Zwittergent 3-14, and then is applied to the column, which is equilibrated with the acetate buffer.

The column is washed with the acetate buffer until the absorbance at 280 nm is stabilized (about 3 column volumes are required). Proteins are

eluted with a 0-0.5 M NaCl gradient in acetate buffer. The fraction eluted at 0.15 M NaCl is enriched with the 76 kDa protein.

The 32 kDa protein (GHPO 1360) can also be purified as follows. Membrane fraction C4 is solubilized in 50 mM $NaCO_3$ buffer (pH 9.5) at room temperature for 30 minutes under stirring. The suspension is then centrifuged at 200,000 x g for 30 minutes at 4°C. This allows the 32 and 35 kDa proteins to be separated, since the 35 kDa protein is insoluble in the $NaCO_3$ buffer. The supernatant is dialyzed against 50 mM $NaPO_4$ buffer (pH 7.0), and then is applied to an SP-Sepharose column, which is equilibrated with the $NaPO_4$ buffer. The column is washed with the $NaPO_4$ buffer, and then an 0-0.5 M NaCl gradient is applied to the column. The fraction eluted between 0.26 and 0.31 M contains the 32 kDa protein.

The 50 kDa protein can also be purified as follows. Membrane fraction C4 is solubilized in 50 mM $NaCO_3$ buffer (pH 9.5) at room temperature for 30 minutes while stirring. The suspension is then centrifuged at 200,000 x g for 30 minutes at 4°C. This allows the 50 and 47 kDa proteins to be separated, since the 47 kDa protein is insoluble in the $NaCO_3$ buffer. The supernatant is dialyzed against 50 mM $NaPO_4$ buffer (pH 7.0).

A 40 ml Q-Sepharose column (diameter: 2.5 cm; height: 8 cm) is prepared according to the manufacturer's instructions (Pharmacia), washed, and equilibrated with buffer B (pH 9.5) (50 mM $NaCO_3$, 100 µM Pefabloc, and 0.1% Zwittergent 3-14).

The chromatography is monitored by UV detection at 280 nm at the column exit. One hundred and forty mg of protein solubilized as is described above are applied to the column, which is then washed with buffer B until the absorbance at 280 nm is stabilized. The proteins are eluted with a 0.1-0.5 M NaCl gradient in buffer B (10 fold V_T), which is followed by washing in

buffer B containing 0.5, and then 1, M NaCl (2 fold V_T). The fractions are recovered, analyzed by SDS-PAGE, and pooled according to their electrophoretic profiles.

Fraction 9, which corresponds to the beginning of the washing at 1 M NaCl and contains acidic proteins, is further purified as follows. A 10 ml DEAE Sepharose column (diameter: 1.5 cm, height: 5 cm) is prepared according to the manufacturer's instructions (Pharmacia) (up to 10 mg protein/ml of gel). The column is washed and equilibrated with buffer B. Chromatography is monitored as is described above.

Fraction 9 is dialyzed against buffer B and contains about 10 mg protein. Fraction 9 is applied to the DEAE-Sepharose column. The column is washed with buffer B until the absorbance at 280 nm is stabilized. The proteins are eluted with a 0-0.5 M NaCl gradient in buffer B (10 fold V_T), followed by washing in buffer B, containing 1 M NaCl (2 fold V_T). Fractions are recovered and analyzed by SDS-PAGE. The 50 kDa protein is found in the fractions eluted at 0.3-0.4 M NaCl.

EXAMPLE 2: Identification of genes in the *H. pylori* genome, such as genes encoding the 76 kDa proteins, the 32 kDa protein (GHPO 1360), and the 50 kDa protein (GHPO 750) identification of signal sequences, and primer design for amplification of genes lacking signal sequences

2.A. Creating *H. pylori* genomic databases

The *H. pylori* genome was provided as a text file containing a single contiguous string of nucleotides that had been determined to be 1.76 Megabases in length. The complete genome was split into 17 separate files using the program SPLIT (Creativity in Action), giving rise to 16 contigs, each

containing 100,000 nucleotides, and a 17th contig containing the remaining 76,000 nucleotides. A header was added to each of the 17 files using the format: >hpg0.txt (representing contig 1), .hpg1.txt (representing contig 2), etc. The resulting 17 files, named hpg0 through hpg16, were then copied together to form one file that represented the plus strand of the complete *H. pylori* genome. The constructed database was given the designation "H." A negative strand database of the *H. pylori* genome was created similarly by first creating a reverse complement of the positive strand using the program SeqPup (D.G. Gilbert, Indiana University Biology Department) and then performing the same procedure as described above for the plus strand. This database was given the designation "N."

The regions predicted to encode open reading frames (ORFs) were defined for the complete *H. pylori* genome using the program GENEMARK™ (Borodovsky *et al.*, Comp. Chem. 17:123, 1993). A database was created from a text file containing an annotated version of all ORFs predicted to be encoded by the *H. pylori* genome for both the plus and minus strands, and was given the designation "O." Each ORF was assigned a number indicating its location on the genome and its position relative to other genes. No manipulation of the text file was required.

2.B. Searching the *H. pylori* databases

The databases constructed as is described above were searched using the program FASTA (Pearson *et al.*, Proc. Natl. Acad. Sci. USA 85:2444-2448, 1988). FASTA was used for searching either a DNA sequence against either of the gene databases ("H" and/or "N"), or a peptide sequence against the ORF library ("O"). TFASTX was used to search a peptide sequence against all possible reading frames of a DNA database ("H" and/or "N" libraries).

Potential frameshifts also being resolved, FASTX was used for searching the translated reading frames of a DNA sequence against either a DNA database, or a peptide sequence against the protein database.

2.C. Isolation of DNA sequences from the *H. pylori* genome

The FASTA searches against the constructed DNA databases identified exact nucleotide coordinates on one or more of the isolated contigs, and therefore the location of the target DNA. Once the exact location of the target sequence was known, the contig identified to carry the gene was exported into the software package MapDraw (DNAStar, Inc.) and the gene was isolated. Gene sequences with flanking DNA was then excised and copied into the EditSeq. Software package (DNAStar, Inc.) for further analysis.

2.D. Identification of signal sequences

The deduced protein encoded by a target gene sequence is analyzed using the PROTEAN software package (DNAStar, Inc.). This analysis predicts those areas of the protein that are hydrophobic by using the Kyte-Doolittle algorithm, and identifies any potential polar residues preceding the hydrophobic core region, which is typical for many signal sequences. For confirmation, the target protein is then searched against a PROSITE database (DNAStar, Inc.) consisting of motifs and signatures. Characteristic of many signal sequences and hydrophobic regions in general, is the identification of predicted prokaryotic lipid attachment sites. Where confirmation between the two approaches is apparent at the N-terminus of any protein, putative cleavage sites are sought. Specifically, this includes the presence of either an Alanine (A), Serine (S), or Glycine (G) residue immediately after the core hydrophobic

region. In the case of lipoproteins, a Cysteine (C) residue would be identified as the +1 residue, post-cleavage.

2.E. Rational design of PCR primers based on the identification of signal sequences

In order to clone gene sequences as N-terminus translational fusions for the generation of recombinant proteins with N-terminal Histidine tags, the gene sequence that specifies the signal sequence is omitted. The 5'-end of the gene-specific portion of the N-terminal primer is designed to start at the first codon beyond the cleavage site. In the case of lipoproteins, the 5'-end of the N-terminal primer begins at the second codon, immediately after the modifiable residue at position +1 post-cleavage. The omission of the signal sequence from the recombinant allows for one-step purification, and potential problems associated with insertion of signal sequences in the membrane of the host strain carrying the hybrid construct are avoided.

EXAMPLE 3: Preparation of isolated DNA encoding GHPO 386, GHPO 789, GHPO 1516, GHPO 896, GHPO 1360, and GHPO 750, and production of these proteins as a histidine-tagged fusion proteins

3.A. Preparation of genomic DNA from *Helicobacter pylori*

Helicobacter pylori strain ORV2001, stored in LB medium containing 50% glycerol at -70°C, is grown on Colombia agar containing 7% sheep blood for 48 hours under microaerophilic conditions (8-10% CO_2, 5-7% O_2, and 85-87% N_2). Cells are harvested, washed with PBS (pH 7.2), and DNA is then extracted from the cells using the Rapid Prep Genomic DNA Isolation kit (Pharmacia Biotech).

3.B. PCR amplification

DNA encoding GHPO 386, GHPO 789, GHPO 1516, GHPO 896, GHPO 1360, and GHPO odd numbers), 65, and 67 is amplified from genomic DNA, as can be prepared as is described above, by the Polymerase Chain Reaction (PCR) using the following primers:

GHPO 386:

N-terminal primer:

5'-CTGAATTCGATTTCAAGGAGAAAACATGAAA-3' (SEQ ID NO:59); and

C-terminal primer:

5'-CCGCTCGAGTTAGTAAGCGAACACATAATT-3' (SEQ ID NO:60).

GHPO 789:

N-terminal primer:

5'-CGCGGATCCGAATCCAATTTAATCCAAAAAGG-3' (SEQ ID NO:61); and

C-terminal primer:

5'-CCGCTCGAGTTAGTAAGCGAACACATAGTTCAA-3' (SEQ ID NO:62).

GHPO 1516:

N-terminal primer:

5'-CGCGGATCCGAATCCAATTTAATCCAAAAAGG-3' (SEQ ID NO:56); and

C-terminal primer:

5'-CCGCTCGAGTTAAGTAAGCGAACACATATTCAA-3' (SEQ ID NO:57).

GHPO 896:

N-terminal primer:

5'-CGCGGATCCGAAGTTTCTTTGTATCAAAG-3' (SEQ ID NO:63); and

C-terminal primer:

5'-CCGCTCGAGTTAGTAAGCAAACACATAATTGTG-3' (SEQ ID NO:64).

GHPO 1360:

N-terminal primer:

5'-CGCGGATCCGAATGAAAAAAAATATCTTAAAT-3' (SEQ ID NO:69); and

C-terminal primer:

5'-CCGCTCGAGTTACTTGTTGATAACAATTTT-3' (SEQ ID NO:70).

GHPO 750:

N-terminal primer:

5'-CGCGGATCCGAATGGCAAAAGAAAAGTTTAAC-3' (SEQ ID NO:71); and

C-terminal primer:

5'-CCGCTCGAGTTATTCAATAATATTGCTCAC-3' (SEQ ID NO:72).

GHPO 711:

N-terminal primer:

5'-GGGAATTCAAAAAAACGAAAAAAACG-3' (SEQ ID NO:83); and

C-terminal primer:

5'-CCCCTCGAGTTAATAGGCAAACAC-3' (SEQ ID NO:84).

The N-terminal and C-terminal primers for each clone both include a 5' clamp and a restriction enzyme recognition sequence for cloning purposes (*Bam*HI (GGATCC) and *Xho*I (CTCGAG) recognition sequences).

Amplification of gene-specific DNA is carried out using a heat-stable DNA Polymerase (*e.g.*, Thermalase DNA Polymerase (Amresco)) according to the manufacturer's instructions. The reaction mixture, which is brought to a

final volume of 100 µl with distilled water, is as follows:

dNTPs mix	200 µM
10x ThermoPol buffer	10 µl
primers	300 nM each
DNA template	50 ng
DNA polymerase	2 units

Appropriate amplification reaction conditions can readily be determined by one skilled in the art. In the present case, the following conditions were used. For GHPO 386 and GHPO 789, in a reaction containing Taq DNA polymerase (Appligene), a denaturing step was carried out at 95°C for 30 seconds, followed by an annealing step at 50°C for one minute, and an extension step at 72°C for 2 minutes and 30 seconds. Twenty five cycles were carried out. For GHPO 896, in a reaction containing Taq DNA polymerase, a denaturing step was carried out at 97°C for 30 seconds, followed by an annealing step at 50°C for one minute, and an extension step at 72°C for 2 minutes and 30 seconds. Twenty five cycles were carried out. The same reaction conditions were used for GHPO 1516 as GHPO 896, except that Vent DNA polymerase was used for clone GHPO 1516, instead of Taq DNA polymerase, and the annealing temperature was 55°C. For GHPO 1360 and GHPO 750, Thermalase DNA polymerase was used. A denaturing step was carried out at 95°C for 30 seconds, followed by an annealing step at 55°C for one minute, and an extension step at 72°C for 2 minutes. Thirty cycles were carried out. For GHPO 711, Vent DNA polymerase was used. A denaturing step was carried out at 94°C for 30 seconds, followed by an annealing step at 50°C for 30 seconds, and an extension step at 72°C for 1 minute. Twenty five cycles were carried out.

3.C. Transformation and selection of transformants

A single PCR product is thus amplified and is then digested at 37°C for 2 hours with *BamH*I and *Xho*I concurrently in a 20 µl reaction volume. The digested product is ligated to similarly cleaved pET28a (Novagen) that is dephosphorylated prior to the ligation by treatment with Calf Intestinal Alkaline Phosphatase (CIP). The gene fusion constructed in this manner allows one-step affinity purification of the resulting fusion protein because of the presence of histidine residues at the N-terminus of the fusion protein, which are encoded by the vector.

The ligation reaction (20 µl) is carried out at 14°C overnight and then is used to transform 100 µl fresh *E. coli* XL1-blue competent cells (Novagen). The cells are incubated on ice for 2 hours, then heat-shocked at 42°C for 30 seconds, and returned to ice for 90 seconds. The samples are then added to 1 ml LB broth in the absence of selection and grown at 37°C for 2 hours. The cells are then plated out on LB agar containing kanamycin (50 µg/ml) at a 10x and neat dilution and incubated overnight at 37°C. The following day, 50 colonies are picked onto secondary plates and incubated at 37°C overnight.

Five colonies are picked into 3 ml LB broth supplemented with kanamycin (100 µg/ml) and are grown overnight at 37°C. Plasmid DNA is extracted using the Quiagen mini-prep. method and is quantitated by agarose gel electrophoresis.

PCR is performed with the gene-specific primers under the conditions stated above and transformant DNA is confirmed to contain the desired insert. If PCR-positive, one of the five plasmid DNA samples (500 ng) extracted from the *E. coli* XL1-blue cells is used to transform competent BL21 (λDE3) *E. coli* competent cells (Novagen; as described previously). Transformants (10) are

picked onto selective kanamycin (50 μg/ml) containing LB agar plates and stored as a research stock in LB containing 50% glycerol.

3.D. Purification of recombinant proteins

One ml of frozen glycerol stock prepared as described in 3.C. is used to inoculate 50 ml of LB medium containing 25 μg/ml of kanamycin in a 250 ml Erlenmeyer flask. The flask is incubated at 37°C for 2 hours or until the absorbance at 600 nm (OD_{600}) reaches 0.4-1.0. The culture is stopped from growing by placing the flask at 4°C overnight. The following day, 10 ml of the overnight culture are used to inoculate 240 ml LB medium containing kanamycin (25 μg/ml), with the initial OD_{600} about 0.02-0.04. Four flasks are inoculated for each ORF.

The cells are grown to an OD_{600} of 1.0 (about 2 hours at 37°C), a 1 ml sample is harvested by centrifugation, and the sample is analyzed by SDS-PAGE to detect any leaky expression. The remaining culture is induced with 1 mM IPTG and the induced cultures are grown for an additional 2 hours at 37°C.

The final OD_{600} is taken and the cells are harvested by centrifugation at 5,000 x g for 15 minutes at 4°C. The supernatant is discarded and the pellets are resuspended in 50 mM Tris-HCl (pH 8.0), 2 mM EDTA. Two hundred and fifty ml of buffer are used for a 1 L culture and the cells are recovered by centrifugation at 12,000 x g for 20 minutes. The supernatant is discarded and the pellets are stored at -45°C.

3. E. Protein purification

Pellets obtained from 3.D. are thawed and resuspended in 95 ml of 50 mM Tris-HCl (pH 8.0). Pefabloc and lysozyme are added to final

concentrations of 100 μM and 100 μg/ml, respectively. The mixture is homogenized with magnetic stirring at 5°C for 30 minutes. Benzonase (Merck) is added at a 1 U/ml final concentration, in the presence of 10 mM $MgCl_2$, to ensure total digestion of the DNA. The suspension is sonicated (Branson Sonifier 450) for 3 cycles of 2 minutes each at maximum output. The homogenate is spun in a centrifuge at 19,000 x g for 15 minutes and both the supernatant and the pellet are analyzed by SDS-PAGE to detect the cellular location of the target protein in the soluble or insoluble fractions, as is described further below.

3.E.1. Soluble fraction

If the target protein is produced in a soluble form (*i.e.*, in the supernatant obtained in 3.E.) NaCl and imidazole are added to the supernatant to final concentrations of 50 mM Tris-HCl (pH 8.0), 0.5 M NaCl, and 10 mM imidazole (buffer A). The mixture is filtered through a 0.45 μm membrane and loaded onto an IMAC column (Pharmacia HiTrap chelating Sepharose; 1 ml) that has been charged with nickel ions according to the manufacturer's recommendations. After loading, the column is washed with 50 column volumes of buffer A and the recombinant target protein is eluted with 5 ml of buffer B (50 mM Tris-HCl (pH 8.0), 0.5 M NaCl, 500 mM imidazole).

The elution profile is monitored by measuring the absorbance of the fractions at 280 nm. Fractions corresponding to the protein peak are pooled, dialyzed against PBS containing 0.5 M arginine, filtered through a 0.22 μm membrane, and stored at -45°C.

3.E.2. Insoluble fraction

If the target protein is expressed in the insoluble fraction (pellets obtained from 3.E.), purification is conducted under denaturing conditions. NaCl, imidazole, and urea are added to the resuspended pellet to final

concentrations of 50 mM Tris-HCl (pH 8.0), 0.5 M NaCl, 10 mM imidazole, and 6 M urea (buffer C). After complete solubilization, the mixture is filtered through a 0.45 μm membrane and loaded onto an IMAC column.

The purification procedures on the IMAC column are the same as described in 3.E.1., except that 6 M urea is included in all buffers used and 10 column volumes of buffer C are used to wash the column after protein loading, instead of 50 column volumes.

The protein fractions eluted from the IMAC column with buffer D (buffer C containing 500 mM imidazole) are pooled. Arginine is added to the solution to final concentration of 0.5 M and the mixture is dialyzed against PBS containing 0.5 M arginine and various concentrations of urea (4 M, 3 M, 2 M, 1 M, and 0.5 M) to progressively decrease the concentration of urea. The final dialysate is filtered through a 0.22 μm membrane and stored at -45°C.

Alternatively, when the above purification process is not as efficient as it should be, two other processes may be used as follows. A first alternative involves the use of a mild denaturant, N-octyl glucoside (NOG). Briefly, a pellet obtained in 3.E. is homogenized in 5 mM imidazole, 500 mM sodium chloride, 20 mM Tris-HCl (pH 7.9) by microfluidization at a pressure of 15,000 psi and is clarified by centrifugation at 4,000-5,000 x g. The pellet is recovered, resuspended in 50 mM $NaPO_4$ (pH 7.5) containing 1-2% weight /volume NOG, and homogenized. The NOG-soluble impurities are removed by centrifugation. The pellet is extracted once more by repeating the preceding extraction step. The pellet is dissolved in 8 M urea, 50 mM Tris (pH 8.0). The urea-solubilized protein is diluted with an equal volume of 2 M arginine, 50 mM Tris (pH 8.0), and is dialyzed against 1 M arginine for 24-48 hours to remove the urea. The final dialysate is filtered through a 0.22 μm membrane and stored at -45°C.

A second alternative involves the use of a strong denaturant, such as guanidine hydrochloride. Briefly, a pellet obtained in 3.E. is homogenized in 5 mM imidazole, 500 mM sodium chloride, 20 mM Tris-HCl (pH 7.9) by microfluidization at a pressure of 15,000 psi and clarified by centrifugation at 4,000-5,000 x g. The pellet is recovered, resuspended in 6 M guanidine hydrochloride, and passed through an IMAC column charged with Ni^{++}. The bound antigen is eluted with 8 M urea (pH 8.5). Beta-mercaptoethanol is added to the eluted protein to a final concentration of 1 mM, then the eluted protein is passed through a Sephadex G-25 column equilibrated in 0.1 M acetic acid. Protein eluted from the column is slowly added to 4 volumes of 50 mM phosphate buffer (pH 7.0). The protein remains in solution.

3.F. Evaluation of the protective activity of the purified protein

A protection test is described above that was carried out for testing the protective activity of the purified, native proteins. This test can also be used for testing the protective efficacy of recombinant proteins. Alternatively, the following test can be used.

Groups of 10 OF1 mice (IFFA Credo) are immunized rectally with 25 µg of the purified recombinant protein, admixed with 1 µg of cholera toxin (Berna) in physiological buffer. Mice are immunized on days 0, 7, 14, and 21. Fourteen days after the last immunization, the mice are challenged with *H. pylori* strain ORV2001 grown in liquid media (the cells are grown on agar plates, as described in 1.A., and, after harvest, the cells are resuspended in Brucella broth; the flasks are then incubated overnight at 37°C). Fourteen days after challenge, the mice are sacrificed and their stomachs are removed. The amount of *H. pylori* is determined by measuring the urease activity in the stomach and by culture.

3.G. Production of monospecific polyclonal antibodies

3.G.1. Hyperimmune rabbit antiserum

New Zealand rabbits are injected both subcutaneously and intramuscularly with 100 µg of a purified fusion polypeptide, as obtained in 3.E.1. or 3.E.2., in the presence of Freund's complete adjuvant and in a total volume of approximately 2 ml. Twenty one and 42 days after the initial injection, booster doses, which are identical to priming doses, except that Freund's incomplete adjuvant is used, are administered in the same way. Fifteen days after the last injection, animal serum is recovered, decomplemented, and filtered through a 0.45 µm membrane.

3.G.2. Mouse hyperimmune ascites fluid

Ten mice are injected subcutaneously with 10-50 µg of a purified fusion polypeptide, as obtained in 3.E.1. or 3.E.2., in the presence of Freund's complete adjuvant and in a volume of approximately 200 µl. Seven and 14 days after the initial injection, booster doses, which are identical to the priming doses, except that Freund's incomplete adjuvant is used, are administered in the same way. Twenty one and 28 days after the initial infection, mice receive 50 µg of the antigen alone intraperitoneally. On day 21, mice are also injected intraperitoneally with sarcoma 180/TG cells CM26684 (Lennette *et al.*, *Diagnostic Procedures for Viral, Rickettsial, and Chlamydial Infections*, 5th Ed., Washington DC, American Public Health Association, 1979). Ascites fluid is collected 10-13 days after the last injection.

EXAMPLE 4: Methods for producing transcriptional fusions lacking His-tags

Methods for amplification and cloning of DNA encoding the polypeptides of the invention as transcriptional fusions lacking His-tags are

described as follows. Two PCR primers for each clone are designed based upon the sequences of the polynucleotides that encode them (SEQ ID NOs:1-21 (odd numbers), 65, and 67). These primers can be used to amplify DNA encoding the polypeptides of the invention from any *Helicobacter pylori* strain, including, for example, ORV2001 and the *H. pylori* strain deposited with the American Type Culture Collection (ATCC, Rockville, Maryland) as ATCC number 43579, as well as from other *Helicobacter* species.

The N-terminal primers are designed to include the ribosome binding site of the target gene, the ATG start site, the signal sequence (if any), and the cleavage site. The N-terminal primers can include a 5' clamp and restriction endonuclease recognition site, such as that for *BamHI* (GGATCC), which facilitates subsequent cloning. Similarly, the C-terminal primers can include a restriction endonuclease recognition site, such as that for *XhoI* (CTCGAG), which can be used in subsequent cloning, and a TAA stop codon. Specific primers that can be used are listed above.

Amplification of a genes encoding the polypeptides of the invention can be carried out using Vent DNA polymerase (New England Biolabs) or Taq DNA polymerase (Appligene) under the conditions described above in Example 3. Alternatively, Thermalase DNA polymerase or Pwo DNA polymerase (Boehringer Mannheim) can be used, according to instructions provided by the manufacturers.

A single PCR product for each clone is amplified and can be cloned into *BamHI-XhoI* cleaved pET24, resulting in construction of transcriptional fusions that permit expression of the proteins without His-tags. The expressed products can be purified as denatured proteins that are refolded by dialysis into 1 M arginine.

Cloning into pET 24 allows transcription of genes from the T7 promoter, which is supplied by the vector, but relies upon binding of the RNA-specific DNA polymerase to the intrinsic ribosome binding site of the genes, and thereby expression of the complete ORF. The amplification, digestion, and cloning protocols are as described above for constructing translational fusions.

EXAMPLE 5: Purification of the polypeptides of the invention by immunoaffinity

5.A. Purification of specific IgGs

An immune serum, as prepared as is described in section 3.G., is applied to a protein A Sepharose Fast Flow column (Pharmacia) equilibrated in 100 mM Tris-HCl (pH 8.0). The resin is washed by applying 10 column volumes of 100 mM Tris-HCl and 10 volumes of 10 mM Tris-HCl (pH 8.0) to the column. IgG antibodies are eluted with 0.1 M glycine buffer (pH 3.0) and are collected in 5 ml fractions to each of which is added 0.25 ml 1 M Tris-HCl (pH 8.0). The optical density of the eluate is measured at 280 nm and the fractions containing the IgG antibodies are pooled, dialyzed against 50 mM Tris-HCl (pH 8.0), and, if necessary, stored frozen at -70°C.

5.B. Preparation of the column

An appropriate amount of CNBr-activated Sepharose 4B gel (1 g of dried gel provides for approximately 3.5 ml of hydrated gel; gel capacity is from 5 to 10 mg coupled IgG/ml of gel) manufactured by Pharmacia (17-0430-01) is suspended in 1 mM HCl buffer and washed using a buchner by adding small quantities of 1 mM HCl buffer. The total volume of buffer is 200 ml per gram of gel.

Purified IgG antibodies are dialyzed for 4 hours at 20±5°C against 50 volumes of 500 mM sodium phosphate buffer (pH 7.5). The antibodies are then diluted in 500 mM phosphate buffer (pH 7.5) to a final concentration of 3 mg/ml.

IgG antibodies are mixed with the gel overnight at 5±3°C. The gel is packed into a chromatography column and is washed with 2 column volumes of 500 mM phosphate buffer (pH 7.5), and 1 column volume of 50 mM sodium phosphate buffer, containing 500 mM NaCl (pH 7.5). The gel is then transferred to a tube, mixed with 100 mM ethanolamine (pH 7.5) for 4 hours at room temperature, and washed twice with 2 column volumes of PBS. The gel is then stored in 1/10,000 PBS/merthiolate. The amount of IgG antibodies coupled to the gel is determined by measuring the optical density (OD) at 280 nm of the IgG solution and the direct eluate, plus washings.

5.C. Adsorption and elution of the antigen

An antigen solution in 50 mM Tris-HCl (pH 8.0), 2 mM EDTA, for example, the supernatant obtained in 3.E.1. or the solubilized pellet obtained in 3.E.2., after centrifugation and filtration through a 0.45 µm membrane, is applied to a column equilibrated with 50 mM Tris-HCl (pH 8.0), 2 mM EDTA, at a flow rate of about 10 ml/hour. The column is then washed with 20 volumes of 50 mM Tris-HCl (pH 8.0), 2 mM EDTA. Alternatively, adsorption can be achieved by mixing overnight at 5±3°C.

The adsorbed gel is washed with 2 to 6 volumes of 10 mM sodium phosphate buffer (pH 6.8) and the antigen is eluted with 100 mM glycine-buffer (pH 2.5). The eluate is recovered in 3 ml fractions, to each of which is added 150 µl of 1 M sodium phosphate buffer (pH 8.0). Absorption is measured at

280 nm for each fraction; those fractions containing the antigen are pooled and stored at

-20°C.

EXAMPLE 6: The GHPO 1360 polypeptide is useful as a serodiagnostic tool for *H. pylori* infection

The reactivity of patient sera against *H. pylori* proteins was analyzed by immunoblot technique. Briefly, total lysate of *H. pylori* strain ORV2001 was subjected to SDS-PAGE electrophoresis (BioRad protean II system) on a 12.5% gel. Proteins were electrotransferred onto a nitrocellulose paper for immunoblot assay. After blocking, the nitrocellulose paper was incubated with patient sera (1:500 diluted in blocking buffer) for one hour at room temperature, washed, and further incubated with peroxidase-conjugated goat anti-human IgG. The positive bands were revealed by incubation with the appropriate substrates. The results showed that the *H. pylori*-positive ulcer patient sera react specifically with proteins having molecular weights between 50 and 60 kDa and about 30 to 35 kDa. To identify the nature of these proteins, the reactivities of the patient sera were analyzed by immunoblot assay against purified proteins with similar molecular weights: urease (67 kDa and 30 kDa), catalase (54 kDa), heat-shock protein B (60 kDa), and the GHPO 1360 polypeptide (32 kDa) expressed and purified as described in Example 5. All patient sera showed strong reactivity against the GHPO 1360 polypeptide, but the reactivities against other purified proteins were quite variable. These results show that the GHPO 1360 polypeptide is a useful antigen for use in diagnosis of *H. pylori* infection.

Other embodiments are within the following claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION

(i) APPLICANT: MERIEUX ORAVAX SOCIETE EN NOM COLLECTIF PASTEUR MERIEUX SERUMS ET VACCINS S.A., ET AL.

(ii) TITLE OF THE INVENTION: 76 kDa, 30 kDa, and 50 kDa Helicobacter Polypeptides and Corresponding Polynucleotide Molecules

(iii) NUMBER OF SEQUENCES: 84

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Clark & Elbing LLP
(B) STREET: 176 Federal Street
(C) CITY: Boston
(D) STATE: MA
(E) COUNTRY: USA
(F) ZIP: 02110

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Diskette
(B) COMPUTER: IBM Compatible
(C) OPERATING SYSTEM: DOS
(D) SOFTWARE: FastSEQ for Windows Version 2.0

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: PCT/US98/----
(B) FILING DATE: 31-MAR-98
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: 08/834,666
(B) FILING DATE: 01-APR-1997

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: 08/831,310
(B) FILING DATE: 01-APR-1997

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Clark, Paul T.
(B) REGISTRATION NUMBER: 30,162
(C) REFERENCE/DOCKET NUMBER: 06132/037WO1

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 617-428-0200
(B) TELEFAX: 617-428-7045
(C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2798 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 328...2451
(D) OTHER INFORMATION:

(A) NAME/KEY: Signal Sequence
(B) LOCATION: 328...385
(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
TGGTCCTGGC ATTCCGAGGT TCGAATCCTT GCACCCCAGC CATTTTTCCT TATTTTTTGG     60
CGCGGAGTAG AGCAGTCCGG TAGCTCGTTG GGCTCATAAC CCAAAGGTCA GTGGTTCAAA    120
TCCATTCTCC GCAACCAATC CTTTAAACCA CACCACCACC AAACGAACCA AACGAAACAA    180
AAAGCATCAA AATCAAAAAA ATGACAAAAT TTTTAAGAAA ATGACAAAAA AAAAAAAAAC    240
GATTTTATGC TATATTAACG AAATCTTGTG ATAAGATCTT ATTCTTTTAA AAGACTTATC    300
TAACCATTTT AATTTCAAGG AGAAAAC ATG AAA AAA ACC CTT TTA CTC TCT CTC    354
                              Met Lys Lys Thr Leu Leu Leu Ser Leu
                                                  -15

TCT CTC TCT CTC TCG TTT TTG CTC CAC GCT GAA GAC GAC GGC TTT TAC     402
Ser Leu Ser Leu Ser Phe Leu Leu His Ala Glu Asp Asp Gly Phe Tyr
-10                 -5                   1               5

ACA AGC GTG GGC TAT CAA ATC GGT GAA GCC GCT CAA ATG GTG AAA AAC     450
Thr Ser Val Gly Tyr Gln Ile Gly Glu Ala Ala Gln Met Val Lys Asn
            10                  15                  20

ACC AAA GGC ATT CAA GAG CTT TCA GAC AAT TAT GAA AAG CTG AAC AAT     498
Thr Lys Gly Ile Gln Glu Leu Ser Asp Asn Tyr Glu Lys Leu Asn Asn
        25                  30                  35

CTT TTG AAT AAT TAC AGC ACC CTA AAC ACC CTT ATC AAA TTG TCC GCT     546
Leu Leu Asn Asn Tyr Ser Thr Leu Asn Thr Leu Ile Lys Leu Ser Ala
    40                  45                  50

GAT CCG AGC GCG ATT AAC GAC GCA AGG GAT AAT CTA GGC TCA AGC TCT     594
Asp Pro Ser Ala Ile Asn Asp Ala Arg Asp Asn Leu Gly Ser Ser Ser
55                  60                  65                  70

AGG AAT TTG CTT GAT GTC AAA ACC AAT TCC CCC GCG TAT CAA GCC GTG     642
Arg Asn Leu Leu Asp Val Lys Thr Asn Ser Pro Ala Tyr Gln Ala Val
                75                  80                  85

CTT TTA GCA CTC AAT GCT GCA GTG GGG TTG TGG CAA GTT ACA AGC TAC     690
```

```
Leu Leu Ala Leu Asn Ala Ala Val Gly Leu Trp Gln Val Thr Ser Tyr
            90                  95                  100

GCT TTT ACT GCT TGT GGT CCT GGC AGT AAC GAG AAT GCG AAT GGA GGG      738
Ala Phe Thr Ala Cys Gly Pro Gly Ser Asn Glu Asn Ala Asn Gly Gly
        105                 110                 115

ATC CAA ACT TTT AAT AAT GTG CCA GGA CAA GAT ACG ACG ACC ATC ACT      786
Ile Gln Thr Phe Asn Asn Val Pro Gly Gln Asp Thr Thr Thr Ile Thr
    120                 125                 130

TGC AAT TCG TAT TAT GAG CCA GGA CAT GGT GGG CCT ATA TCC ACT GCA      834
Cys Asn Ser Tyr Tyr Glu Pro Gly His Gly Gly Pro Ile Ser Thr Ala
135                 140                 145                 150

AAT TAT GCG AAA ATC AAT CAA GCC TAT CAA ATC ATC CAA AAG GCT TTG      882
Asn Tyr Ala Lys Ile Asn Gln Ala Tyr Gln Ile Ile Gln Lys Ala Leu
                155                 160                 165

ACA GCC AAT GGA GCT AAT GGA GAT GGG GTC CCC GTT TTA AGC AAC ACC      930
Thr Ala Asn Gly Ala Asn Gly Asp Gly Val Pro Val Leu Ser Asn Thr
            170                 175                 180

ACT ACA AAA CTT GAT TTC ACT ATC AAT GGA GAC AAA AGA ACG GGG GGC      978
Thr Thr Lys Leu Asp Phe Thr Ile Asn Gly Asp Lys Arg Thr Gly Gly
        185                 190                 195

AAA CCA AAT ACA CCT GAA AAG TTC CCA TGG AGT GAT GGG AAA TAT ATT     1026
Lys Pro Asn Thr Pro Glu Lys Phe Pro Trp Ser Asp Gly Lys Tyr Ile
    200                 205                 210

CAC ACC CAA TGG ATT AAC ACA ATA GTA ACA CCA ACA GAA ACA AAT ATC     1074
His Thr Gln Trp Ile Asn Thr Ile Val Thr Pro Thr Glu Thr Asn Ile
215                 220                 225                 230

AAC ACA GAA AAT AAC GCT CAA GAG CTT TTA AAA CAA GCG AGC ATC ATT     1122
Asn Thr Glu Asn Asn Ala Gln Glu Leu Leu Lys Gln Ala Ser Ile Ile
                235                 240                 245

ATC ACT ACC CTA AAT GAG GCA TGC CCA AAC TTC CAA AAT GGT GGT AGA     1170
Ile Thr Thr Leu Asn Glu Ala Cys Pro Asn Phe Gln Asn Gly Gly Arg
            250                 255                 260

AGT TAT TGG CAA GGG ATA AGC GGC AAT GGG ACA ATG TGC GGG ATG TTT     1218
Ser Tyr Trp Gln Gly Ile Ser Gly Asn Gly Thr Met Cys Gly Met Phe
        265                 270                 275

AAG AAT GAA ATC AGC GCG ATC CAA GGC ATG ATC GCT AAC GCT CAA GAA     1266
Lys Asn Glu Ile Ser Ala Ile Gln Gly Met Ile Ala Asn Ala Gln Glu
    280                 285                 290

GCT GTC GCG CAA AGC AAA ATC GTT AGT GAA AAC GCG CAA AAT CAA AAC     1314
Ala Val Ala Gln Ser Lys Ile Val Ser Glu Asn Ala Gln Asn Gln Asn
295                 300                 305                 310

AAC TTG GAT ACT GGA AAA CCA TTC AAC CCT TAC ACG GAC GCC AGC TTT     1362
Asn Leu Asp Thr Gly Lys Pro Phe Asn Pro Tyr Thr Asp Ala Ser Phe
```

```
                315                 320                 325

GCG CAA AGC ATG CTC AAA AAC GCT CAA GCG CAA GCA GAG ATT TTA AAC    1410
Ala Gln Ser Met Leu Lys Asn Ala Gln Ala Gln Ala Glu Ile Leu Asn
            330                 335                 340

CAA GCC GAA CAA GTA GTA AAA AAC TTT GAA AAA ATC CCT ACA GCC TTT    1458
Gln Ala Glu Gln Val Val Lys Asn Phe Glu Lys Ile Pro Thr Ala Phe
        345                 350                 355

GTA TCA GAC TCT TTA GGG GTG TGT TAT GAA GTG CAA GGG GGT GAG CGT    1506
Val Ser Asp Ser Leu Gly Val Cys Tyr Glu Val Gln Gly Gly Glu Arg
    360                 365                 370

AGG GGC ACC AAT CCA GGT CAG GTA ACT TCT AAC ACT TGG GGA GCC GGT    1554
Arg Gly Thr Asn Pro Gly Gln Val Thr Ser Asn Thr Trp Gly Ala Gly
375                 380                 385                 390

TGC GCG TAT GTG AAA CAA ACC ATA ACG AAT TTA GAC AAC AGC ATC GCT    1602
Cys Ala Tyr Val Lys Gln Thr Ile Thr Asn Leu Asp Asn Ser Ile Ala
                395                 400                 405

CAC TTT GGC ACT CAA GAG CAG CAG ATA CAG CAA GCC GAA AAC ATC GCT    1650
His Phe Gly Thr Gln Glu Gln Gln Ile Gln Gln Ala Glu Asn Ile Ala
            410                 415                 420

GAC ACT CTA GTG AAT TTC AAA TCT AGA TAC AGC GAA TTA GGC AAC ACC    1698
Asp Thr Leu Val Asn Phe Lys Ser Arg Tyr Ser Glu Leu Gly Asn Thr
        425                 430                 435

TAT AAC AGC ATC ACC ACC GCG CTC TCC AAA GTC CCT AAC GCG CAA AGC    1746
Tyr Asn Ser Ile Thr Thr Ala Leu Ser Lys Val Pro Asn Ala Gln Ser
    440                 445                 450

TTG CAA AAC GTG GTG AGC AAA AAG AAT AAC CCC TAT AGC CCT CAA GGC    1794
Leu Gln Asn Val Val Ser Lys Lys Asn Asn Pro Tyr Ser Pro Gln Gly
455                 460                 465                 470

ATA GAG ACC AAT TAC TAC CTC AAT CAA AAT TCT TAC AAC CAA ATC CAA    1842
Ile Glu Thr Asn Tyr Tyr Leu Asn Gln Asn Ser Tyr Asn Gln Ile Gln
                475                 480                 485

ACC ATC AAC CAA GAA CTA GGG CGT AAC CCC TTT AGG AAA GTG GGC ATC    1890
Thr Ile Asn Gln Glu Leu Gly Arg Asn Pro Phe Arg Lys Val Gly Ile
            490                 495                 500

GTC AAT TCT CAA ACC AAC AAT GGT GCC ATG AAT GGG ATC GGT ATT CAG    1938
Val Asn Ser Gln Thr Asn Asn Gly Ala Met Asn Gly Ile Gly Ile Gln
        505                 510                 515

GTG GGC TAT AAG CAA TTC TTT GGC CAA AAA AGA AAA TGG GGC GCT AGG    1986
Val Gly Tyr Lys Gln Phe Phe Gly Gln Lys Arg Lys Trp Gly Ala Arg
    520                 525                 530

TAT TAC GGC TTT TTT GAC TAC AAC CAT GCG TTC ATT AAA TCC AGC TTC    2034
Tyr Tyr Gly Phe Phe Asp Tyr Asn His Ala Phe Ile Lys Ser Ser Phe
535                 540                 545                 550
```

```
TTC AAC TCG GCT TCT GAT GTG TGG ACT TAT GGT TTT GGA GCG GAC GCT    2082
Phe Asn Ser Ala Ser Asp Val Trp Thr Tyr Gly Phe Gly Ala Asp Ala
                555                 560                 565

CTT TAT AAC TTC ATC AAC GAT AAA GCC ACC AAT TTC TTA GGC AAA AAC    2130
Leu Tyr Asn Phe Ile Asn Asp Lys Ala Thr Asn Phe Leu Gly Lys Asn
            570                 575                 580

AAC AAG CTT TCC GTG GGG CTT TTT GGA GGG ATT GCG TTA GCG GGC ACT    2178
Asn Lys Leu Ser Val Gly Leu Phe Gly Gly Ile Ala Leu Ala Gly Thr
        585                 590                 595

TCA TGG CTT AAT TCT GAG TAT GTG AAT TTA GCC ACC GTG AAT AAC GTC    2226
Ser Trp Leu Asn Ser Glu Tyr Val Asn Leu Ala Thr Val Asn Asn Val
    600                 605                 610

TAT AAC GCT AAA ATG AAT GTG GCG AAT TTC CAA TTC TTA TTC AAT ATG    2274
Tyr Asn Ala Lys Met Asn Val Ala Asn Phe Gln Phe Leu Phe Asn Met
615                 620                 625                 630

GGA GTG AGG ATG AAT TTA GCC AGA TCC AAG AAA AAA GGC AGC GAT CAT    2322
Gly Val Arg Met Asn Leu Ala Arg Ser Lys Lys Lys Gly Ser Asp His
                635                 640                 645

GCG GCT CAG CAT GGG ATT GAA CTA GGG CTT AAA ATC CCC ACC ATC AAC    2370
Ala Ala Gln His Gly Ile Glu Leu Gly Leu Lys Ile Pro Thr Ile Asn
            650                 655                 660

ACG AAC TAT TAT TCT TTC ATG GGG GCT GAA CTC AAA TAC AGA AGG CTT    2418
Thr Asn Tyr Tyr Ser Phe Met Gly Ala Glu Leu Lys Tyr Arg Arg Leu
        665                 670                 675

TAT AGC GTG TAT TTG AAT TAT GTG TTC GCT TAC TAAGCTTTTT GTGAAACTCC   2471
Tyr Ser Val Tyr Leu Asn Tyr Val Phe Ala Tyr
    680                 685

CTTTTTAAGG GGTTTTTTTT TGAACTCTCT TTTTAAATTC TCTTTTTAAA GAGATTTCTT   2531
TTTTTTAAGC TTTTTTTTGA ATTCTTTTTT TTGAATTCTT TGTTTTTAAG CTTTTTTTAA   2591
ACCCTTTCGT TTTTAAACTC CCTTTTTTAA GGGATTTCTT TTTTTAAACT CTTTTTTTTT   2651
AAACTCTTTT TTTTAAACCC TCTTTTTTTA AGGGATTTCT TTTTAAAGCT TTTTTGAAGT   2711
CTTTTTTTAA ATTCTTTTTT TGGGGGTTTG ATCTTTCTTT TTGCCAATCC CCACTACTTT   2771
CGCTTTTTAA TCTTTAGGTT TTATTTT                                       2798
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 708 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal
(ix) FEATURE:

(A) NAME/KEY: Signal Sequence
(B) LOCATION: 1...19

(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Lys Lys Thr Leu Leu Leu Ser Leu Ser Leu Ser Leu Ser Phe Leu
                -15                 -10                 -5
Leu His Ala Glu Asp Asp Gly Phe Tyr Thr Ser Val Gly Tyr Gln Ile
            1               5                  10
Gly Glu Ala Ala Gln Met Val Lys Asn Thr Lys Gly Ile Gln Glu Leu
    15                  20                  25
Ser Asp Asn Tyr Glu Lys Leu Asn Asn Leu Leu Asn Asn Tyr Ser Thr
30                  35                  40                  45
Leu Asn Thr Leu Ile Lys Leu Ser Ala Asp Pro Ser Ala Ile Asn Asp
                50                  55                  60
Ala Arg Asp Asn Leu Gly Ser Ser Ser Arg Asn Leu Leu Asp Val Lys
            65                  70                  75
Thr Asn Ser Pro Ala Tyr Gln Ala Val Leu Leu Ala Leu Asn Ala Ala
        80                  85                  90
Val Gly Leu Trp Gln Val Thr Ser Tyr Ala Phe Thr Ala Cys Gly Pro
    95                  100                 105
Gly Ser Asn Glu Asn Ala Asn Gly Gly Ile Gln Thr Phe Asn Asn Val
110                 115                 120                 125
Pro Gly Gln Asp Thr Thr Thr Ile Thr Cys Asn Ser Tyr Tyr Glu Pro
                130                 135                 140
Gly His Gly Gly Pro Ile Ser Thr Ala Asn Tyr Ala Lys Ile Asn Gln
            145                 150                 155
Ala Tyr Gln Ile Ile Gln Lys Ala Leu Thr Ala Asn Gly Ala Asn Gly
        160                 165                 170
Asp Gly Val Pro Val Leu Ser Asn Thr Thr Thr Lys Leu Asp Phe Thr
    175                 180                 185
Ile Asn Gly Asp Lys Arg Thr Gly Gly Lys Pro Asn Thr Pro Glu Lys
190                 195                 200                 205
Phe Pro Trp Ser Asp Gly Lys Tyr Ile His Thr Gln Trp Ile Asn Thr
                210                 215                 220
Ile Val Thr Pro Thr Glu Thr Asn Ile Asn Thr Glu Asn Asn Ala Gln
            225                 230                 235
Glu Leu Leu Lys Gln Ala Ser Ile Ile Ile Thr Thr Leu Asn Glu Ala
        240                 245                 250
Cys Pro Asn Phe Gln Asn Gly Gly Arg Ser Tyr Trp Gln Gly Ile Ser
    255                 260                 265
Gly Asn Gly Thr Met Cys Gly Met Phe Lys Asn Glu Ile Ser Ala Ile
270                 275                 280                 285
Gln Gly Met Ile Ala Asn Ala Gln Glu Ala Val Ala Gln Ser Lys Ile
                290                 295                 300
Val Ser Glu Asn Ala Gln Asn Gln Asn Asn Leu Asp Thr Gly Lys Pro
            305                 310                 315
Phe Asn Pro Tyr Thr Asp Ala Ser Phe Ala Gln Ser Met Leu Lys Asn
        320                 325                 330
Ala Gln Ala Gln Ala Glu Ile Leu Asn Gln Ala Glu Gln Val Val Lys
    335                 340                 345
Asn Phe Glu Lys Ile Pro Thr Ala Phe Val Ser Asp Ser Leu Gly Val
350                 355                 360                 365
Cys Tyr Glu Val Gln Gly Gly Glu Arg Arg Gly Thr Asn Pro Gly Gln
                370                 375                 380
Val Thr Ser Asn Thr Trp Gly Ala Gly Cys Ala Tyr Val Lys Gln Thr
```

```
                385                 390                 395
Ile Thr Asn Leu Asp Asn Ser Ile Ala His Phe Gly Thr Gln Glu Gln
        400                 405                 410
Gln Ile Gln Gln Ala Glu Asn Ile Ala Asp Thr Leu Val Asn Phe Lys
    415                 420                 425
Ser Arg Tyr Ser Glu Leu Gly Asn Thr Tyr Asn Ser Ile Thr Thr Ala
430                 435                 440                 445
Leu Ser Lys Val Pro Asn Ala Gln Ser Leu Gln Asn Val Val Ser Lys
                450                 455                 460
Lys Asn Asn Pro Tyr Ser Pro Gln Gly Ile Glu Thr Asn Tyr Tyr Leu
            465                 470                 475
Asn Gln Asn Ser Tyr Asn Gln Ile Gln Thr Ile Asn Gln Glu Leu Gly
        480                 485                 490
Arg Asn Pro Phe Arg Lys Val Gly Ile Val Asn Ser Gln Thr Asn Asn
    495                 500                 505
Gly Ala Met Asn Gly Ile Gly Ile Gln Val Gly Tyr Lys Gln Phe Phe
510                 515                 520                 525
Gly Gln Lys Arg Lys Trp Gly Ala Arg Tyr Tyr Gly Phe Phe Asp Tyr
                530                 535                 540
Asn His Ala Phe Ile Lys Ser Ser Phe Phe Asn Ser Ala Ser Asp Val
            545                 550                 555
Trp Thr Tyr Gly Phe Gly Ala Asp Ala Leu Tyr Asn Phe Ile Asn Asp
        560                 565                 570
Lys Ala Thr Asn Phe Leu Gly Lys Asn Asn Lys Leu Ser Val Gly Leu
    575                 580                 585
Phe Gly Gly Ile Ala Leu Ala Gly Thr Ser Trp Leu Asn Ser Glu Tyr
590                 595                 600                 605
Val Asn Leu Ala Thr Val Asn Asn Val Tyr Asn Ala Lys Met Asn Val
                610                 615                 620
Ala Asn Phe Gln Phe Leu Phe Asn Met Gly Val Arg Met Asn Leu Ala
            625                 630                 635
Arg Ser Lys Lys Lys Gly Ser Asp His Ala Ala Gln His Gly Ile Glu
        640                 645                 650
Leu Gly Leu Lys Ile Pro Thr Ile Asn Thr Asn Tyr Tyr Ser Phe Met
    655                 660                 665
Gly Ala Glu Leu Lys Tyr Arg Arg Leu Tyr Ser Val Tyr Leu Asn Tyr
670                 675                 680                 685
Val Phe Ala Tyr
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2699 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 199...2397
(D) OTHER INFORMATION:

(A) NAME/KEY: Signal Sequence
(B) LOCATION: 199...259
(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
TAAAATCCAA TTAAAAGCGT TCAAAGGTAA CGCAAAAAAA CAAAAAATGA CGCAATTTTT      60
TCAAAATGAC AAAAAAAAAC GCTTTATGCT ATAATACCCC AAATACATTC TAATAGCAAA     120
TGCGTTCTAA TGCAAATGCA TTCCAATGTA TGAAATCCCT AATACTAAAT CCAATTTAAT     180
CCAAAAAGGA GAAAAAAC ATG AAA AAA CAC ATC CTT TCA TTA GCT TTA GGC       231
                    Met Lys Lys His Ile Leu Ser Leu Ala Leu Gly
                    -20                 -15                 -10

TCG CTT TTA GTT TCC ACT TTG AGC GCT GAA GAC GAC GGC TTT TAC ACA       279
Ser Leu Leu Val Ser Thr Leu Ser Ala Glu Asp Asp Gly Phe Tyr Thr
                -5                   1                   5

AGC GTA GGC TAT CAG ATC GGT GAA GCC GCT CAA ATG GTA ACA AAC ACC       327
Ser Val Gly Tyr Gln Ile Gly Glu Ala Ala Gln Met Val Thr Asn Thr
        10                  15                  20

AAA GGC ATC CAA CAG CTT TCA GAC AAT TAT GAA AAT TTG AAC AAC CTT       375
Lys Gly Ile Gln Gln Leu Ser Asp Asn Tyr Glu Asn Leu Asn Asn Leu
    25                  30                  35

TTA ACG AGA TAC AGC ACC CTA AAC ACC CTT ATC AAA TTG TCC GCT GAT       423
Leu Thr Arg Tyr Ser Thr Leu Asn Thr Leu Ile Lys Leu Ser Ala Asp
40                  45                  50                  55

CCG AGC GCA ATT AAT GCG GTG CGG GAA AAT CTG GGC GCG AGC GCG AAG       471
Pro Ser Ala Ile Asn Ala Val Arg Glu Asn Leu Gly Ala Ser Ala Lys
                60                  65                  70

AAT TTG ATC GGC GAT AAA GCC AAC TCC CCC GCC TAT CAA GCC GTG CTT       519
Asn Leu Ile Gly Asp Lys Ala Asn Ser Pro Ala Tyr Gln Ala Val Leu
            75                  80                  85

TTA GCG ATC AAC GCG GCG GTA GGG TTT TGG AAT GTC GTG GGC TAT GTG       567
Leu Ala Ile Asn Ala Ala Val Gly Phe Trp Asn Val Val Gly Tyr Val
        90                  95                  100

ACG CAA TGT GGG GGT AAC GCC AAT GGT CAA GAA AGC ACC TCT TCA ACC       615
Thr Gln Cys Gly Gly Asn Ala Asn Gly Gln Glu Ser Thr Ser Ser Thr
    105                 110                 115

ACC ATC TTC AAC AAC GAG CCA GGG TAT CGA TCC ACT TCC ATC ACT TGT       663
Thr Ile Phe Asn Asn Glu Pro Gly Tyr Arg Ser Thr Ser Ile Thr Cys
120                 125                 130                 135

TCT TTG AAC GGG CAT AAG CCT GGA TAC TAT GGC CCT ATG AGC ATT GAG       711
Ser Leu Asn Gly His Lys Pro Gly Tyr Tyr Gly Pro Met Ser Ile Glu
                140                 145                 150

AAT TTT AAA AAG CTT AAC GAA GCC TAT CAG ATC CTC CAA ACG GCT TTA       759
Asn Phe Lys Lys Leu Asn Glu Ala Tyr Gln Ile Leu Gln Thr Ala Leu
```

```
                155                 160                 165

AAA AAC GGC TTA CCC GCG CTC AAA GAA AAC AAC GGG AAG GTC AGT GTA     807
Lys Asn Gly Leu Pro Ala Leu Lys Glu Asn Asn Gly Lys Val Ser Val
        170                 175                 180

ACC TAT ACC TAC ACA TGC TCA GGG CAA GGG AAT AAT AAC TGC TCG CCA     855
Thr Tyr Thr Tyr Thr Cys Ser Gly Gln Gly Asn Asn Asn Cys Ser Pro
    185                 190                 195

AGT GTC AAC GGA ACC AAA ACC ACA ACC CAA ACC ATA GAC GGC AAA AGC     903
Ser Val Asn Gly Thr Lys Thr Thr Thr Gln Thr Ile Asp Gly Lys Ser
200                 205                 210                 215

GTA ACC ACC ACG ATC AGT TCA AAA GTG GTT GGT AGC ATC GCT AGT GGC     951
Val Thr Thr Thr Ile Ser Ser Lys Val Val Gly Ser Ile Ala Ser Gly
                220                 225                 230

AAC ACA TCA CAT GTC ATC ACC AAC AAA TTA GAC GGT GTG CCT GAT AGC     999
Asn Thr Ser His Val Ile Thr Asn Lys Leu Asp Gly Val Pro Asp Ser
            235                 240                 245

GCT CAA GCG CTC TTA GCG CAA GCG AGC ACG CTC ATC AAC ACC ATC AAC    1047
Ala Gln Ala Leu Leu Ala Gln Ala Ser Thr Leu Ile Asn Thr Ile Asn
        250                 255                 260

GAA GCA TGC CCG TAT TTC CAT GCT ACT AAT AGT AGT GAG GCT AAC GCC    1095
Glu Ala Cys Pro Tyr Phe His Ala Thr Asn Ser Ser Glu Ala Asn Ala
    265                 270                 275

CCA AAA TTC TCT ACT ACT ACT GGG AAA ATA TGC GGC GCT TTT TCA GAA    1143
Pro Lys Phe Ser Thr Thr Thr Gly Lys Ile Cys Gly Ala Phe Ser Glu
280                 285                 290                 295

GAA ATC AGC GCG ATC CAA AAG ATG ATC ACG GAC GCG CAA GAG CTA GTT    1191
Glu Ile Ser Ala Ile Gln Lys Met Ile Thr Asp Ala Gln Glu Leu Val
                300                 305                 310

AAT CAA ACG AGC GTC ATT AAC AGC AAC GAA CAA TCA ACT CCG GTA GGC    1239
Asn Gln Thr Ser Val Ile Asn Ser Asn Glu Gln Ser Thr Pro Val Gly
            315                 320                 325

AAT AAT AAT GGC AAG CCT TTC AAC CCT TTC ACG GAC GCA AGT TTT GCG    1287
Asn Asn Asn Gly Lys Pro Phe Asn Pro Phe Thr Asp Ala Ser Phe Ala
        330                 335                 340

CAA GGC ATG CTC GCT AAC GCT AGC GCG CAA GCT AAA ATG CTC AAT TTA    1335
Gln Gly Met Leu Ala Asn Ala Ser Ala Gln Ala Lys Met Leu Asn Leu
    345                 350                 355

GCC CAT CAG GTG GGG CAA GCC ATT AAC CCA GAG AAT CTT AGC GAG AAT    1383
Ala His Gln Val Gly Gln Ala Ile Asn Pro Glu Asn Leu Ser Glu Asn
360                 365                 370                 375

TTT AAA AAT TTT GTT ACA GGC TTT TTA GCC ACA TGC AAT AAC AAA TCA    1431
Phe Lys Asn Phe Val Thr Gly Phe Leu Ala Thr Cys Asn Asn Lys Ser
                380                 385                 390
```

```
ACA GCT GGC ACT GGT GGC ACA CAA GGT TCA GCT CCA GGC ACA GTG ACC    1479
Thr Ala Gly Thr Gly Gly Thr Gln Gly Ser Ala Pro Gly Thr Val Thr
            395                 400                 405

ACT CAA ACT TTC GCT TCT GGT TGC GCG TAT GTG GAG CAA ACC CTA ACG    1527
Thr Gln Thr Phe Ala Ser Gly Cys Ala Tyr Val Glu Gln Thr Leu Thr
        410                 415                 420

AAC TTA GGC AAC AGC ATC GCT CAC TTT GGC ACT CAA GAG CAG CAG ATA    1575
Asn Leu Gly Asn Ser Ile Ala His Phe Gly Thr Gln Glu Gln Gln Ile
    425                 430                 435

CAG CAA GCC GAA AAC ATC GCT GAC ACT CTA GTG AAT TTC AAA TCT AGA    1623
Gln Gln Ala Glu Asn Ile Ala Asp Thr Leu Val Asn Phe Lys Ser Arg
440                 445                 450                 455

TAC AGC GAA TTA GGC AAC ACC TAT AAC AGC ATC ACC ACC GCG CTC TCC    1671
Tyr Ser Glu Leu Gly Asn Thr Tyr Asn Ser Ile Thr Thr Ala Leu Ser
                460                 465                 470

AAA GTC CCT AAC GCG CAA AGC TTG CAA AAC GTG GTG AGC AAA AAG AAT    1719
Lys Val Pro Asn Ala Gln Ser Leu Gln Asn Val Val Ser Lys Lys Asn
            475                 480                 485

AAC CCC TAT AGC CCT CAA GGC ATA GAG ACC AAT TAC TAC CTC AAT CAA    1767
Asn Pro Tyr Ser Pro Gln Gly Ile Glu Thr Asn Tyr Tyr Leu Asn Gln
        490                 495                 500

AAT TCT TAC AAC CAA ATC CAA ACC ATC AAC CAA GAA CTA GGG CGT AAC    1815
Asn Ser Tyr Asn Gln Ile Gln Thr Ile Asn Gln Glu Leu Gly Arg Asn
    505                 510                 515

CCC TTT AGG AAA GTG GGC ATC GTC AAT TCT CAA ACC AAC AAT GGT GCC    1863
Pro Phe Arg Lys Val Gly Ile Val Asn Ser Gln Thr Asn Asn Gly Ala
520                 525                 530                 535

ATG AAT GGG ATC GGT ATT CAG GTG GGC TAT AAG CAA TTC TTT GGC CAA    1911
Met Asn Gly Ile Gly Ile Gln Val Gly Tyr Lys Gln Phe Phe Gly Gln
                540                 545                 550

AAA AGA AAA TGG GGC GCT AGG TAT TAC GGC TTT TTT GAT TAC AAC CAT    1959
Lys Arg Lys Trp Gly Ala Arg Tyr Tyr Gly Phe Phe Asp Tyr Asn His
            555                 560                 565

GCG TTC ATC AAA TCC AGC TTT TTC AAC TCG GCT TCT GAC GTG TGG ACT    2007
Ala Phe Ile Lys Ser Ser Phe Phe Asn Ser Ala Ser Asp Val Trp Thr
        570                 575                 580

TAT GGT TTT GGA GCG GAC GCG CTT TAT AAC TTC ATC AAC GAT AAA GCC    2055
Tyr Gly Phe Gly Ala Asp Ala Leu Tyr Asn Phe Ile Asn Asp Lys Ala
    585                 590                 595

ACC AAT TTC TTA GGC AAA AAC AAC AAG CTT TCT TTG GGG CTT TTT GGC    2103
Thr Asn Phe Leu Gly Lys Asn Asn Lys Leu Ser Leu Gly Leu Phe Gly
600                 605                 610                 615

GGG ATT GCG TTA GCG GGC ACT TCA TGG CTC AAT TCT GAG TAC GTG AAT    2151
```

```
Gly Ile Ala Leu Ala Gly Thr Ser Trp Leu Asn Ser Glu Tyr Val Asn
                620                 625                 630

TTA GCC ACC GTG AAT AAC GTC TAT AAC GCT AAA ATG AAT GTG GCG AAT    2199
Leu Ala Thr Val Asn Asn Val Tyr Asn Ala Lys Met Asn Val Ala Asn
           635                 640                 645

TTC CAA TTC TTA TTC AAT ATG GGA GTG AGG ATG AAT TTA GCC AGA TCC    2247
Phe Gln Phe Leu Phe Asn Met Gly Val Arg Met Asn Leu Ala Arg Ser
        650                 655                 660

AAG AAA AAA GGC AGC GAT CAT GCA GCT CAG CAT GGG ATT GAG TTA GGG    2295
Lys Lys Lys Gly Ser Asp His Ala Ala Gln His Gly Ile Glu Leu Gly
    665                 670                 675

CTT AAA ATC CCC ACC ATC AAC ACG AAC TAT TAT TCC TTT ATG GGG GCT    2343
Leu Lys Ile Pro Thr Ile Asn Thr Asn Tyr Tyr Ser Phe Met Gly Ala
680                 685                 690                 695

GAA CTC AAA TAC AGA AGG CTC TAT AGC GTG TAT TTG AAC TAT GTG TTC    2391
Glu Leu Lys Tyr Arg Arg Leu Tyr Ser Val Tyr Leu Asn Tyr Val Phe
                700                 705                 710

GCT TAC TAATGTTTGG CTCTTTGTGA AACTCCCTTT TTAAGGGGTT TTTTTTGAA CT  2449
Ala Tyr

CTCTTTTTAA ATTCTCTTTT TAAAGAGATT TCTTTTTTTT AAGCTTTTTT TTGAATTCTT   2509
TTTTTTTGAA TTCTTTGTTT TTAAGCTTTT TTTAAACCCT TTCGTTTTTA AACTCCCTTT   2569
TTTAAGGGAT TTCTTTTTTT GAACTCCCTT TTTTGAACCC TTTTTTTTAA ACCCTCTTTT   2629
TTTAAGGGGT TTCTTTTTAA AGCTTTTTTG AAGTCTTTTT TTAAATTCTT TTTTTGGGGG   2689
TTTGATCTTT                                                         2699
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 733 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal
(ix) FEATURE:

(A) NAME/KEY: Signal Sequence
(B) LOCATION: 1...20
(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
Met Lys Lys His Ile Leu Ser Leu Ala Leu Gly Ser Leu Leu Val Ser
-20                 -15                 -10                 -5
Thr Leu Ser Ala Glu Asp Asp Gly Phe Tyr Thr Ser Val Gly Tyr Gln
                 1               5                  10
```

Ile Gly Glu Ala Ala Gln Met Val Thr Asn Thr Lys Gly Ile Gln Gln
15 20 25
Leu Ser Asp Asn Tyr Glu Asn Leu Asn Asn Leu Leu Thr Arg Tyr Ser
30 35 40
Thr Leu Asn Thr Leu Ile Lys Leu Ser Ala Asp Pro Ser Ala Ile Asn
45 50 55 60
Ala Val Arg Glu Asn Leu Gly Ala Ser Ala Lys Asn Leu Ile Gly Asp
65 70 75
Lys Ala Asn Ser Pro Ala Tyr Gln Ala Val Leu Leu Ala Ile Asn Ala
80 85 90
Ala Val Gly Phe Trp Asn Val Val Gly Tyr Val Thr Gln Cys Gly Gly
95 100 105
Asn Ala Asn Gly Gln Glu Ser Thr Ser Ser Thr Thr Ile Phe Asn Asn
110 115 120
Glu Pro Gly Tyr Arg Ser Thr Ser Ile Thr Cys Ser Leu Asn Gly His
125 130 135 140
Lys Pro Gly Tyr Tyr Gly Pro Met Ser Ile Glu Asn Phe Lys Lys Leu
145 150 155
Asn Glu Ala Tyr Gln Ile Leu Gln Thr Ala Leu Lys Asn Gly Leu Pro
160 165 170
Ala Leu Lys Glu Asn Asn Gly Lys Val Ser Val Thr Tyr Thr Tyr Thr
175 180 185
Cys Ser Gly Gln Gly Asn Asn Asn Cys Ser Pro Ser Val Asn Gly Thr
190 195 200
Lys Thr Thr Thr Gln Thr Ile Asp Gly Lys Ser Val Thr Thr Thr Ile
205 210 215 220
Ser Ser Lys Val Val Gly Ser Ile Ala Ser Gly Asn Thr Ser His Val
225 230 235
Ile Thr Asn Lys Leu Asp Gly Val Pro Asp Ser Ala Gln Ala Leu Leu
240 245 250
Ala Gln Ala Ser Thr Leu Ile Asn Thr Ile Asn Glu Ala Cys Pro Tyr
255 260 265
Phe His Ala Thr Asn Ser Ser Glu Ala Asn Ala Pro Lys Phe Ser Thr
270 275 280
Thr Thr Gly Lys Ile Cys Gly Ala Phe Ser Glu Glu Ile Ser Ala Ile
285 290 295 300
Gln Lys Met Ile Thr Asp Ala Gln Glu Leu Val Asn Gln Thr Ser Val
305 310 315
Ile Asn Ser Asn Glu Gln Ser Thr Pro Val Gly Asn Asn Asn Gly Lys
320 325 330
Pro Phe Asn Pro Phe Thr Asp Ala Ser Phe Ala Gln Gly Met Leu Ala
335 340 345
Asn Ala Ser Ala Gln Ala Lys Met Leu Asn Leu Ala His Gln Val Gly
350 355 360
Gln Ala Ile Asn Pro Glu Asn Leu Ser Glu Asn Phe Lys Asn Phe Val
365 370 375 380
Thr Gly Phe Leu Ala Thr Cys Asn Asn Lys Ser Thr Ala Gly Thr Gly
385 390 395
Gly Thr Gln Gly Ser Ala Pro Gly Thr Val Thr Thr Gln Thr Phe Ala
400 405 410
Ser Gly Cys Ala Tyr Val Glu Gln Thr Leu Thr Asn Leu Gly Asn Ser
415 420 425
Ile Ala His Phe Gly Thr Gln Glu Gln Gln Ile Gln Gln Ala Glu Asn
430 435 440
Ile Ala Asp Thr Leu Val Asn Phe Lys Ser Arg Tyr Ser Glu Leu Gly
445 450 455 460
Asn Thr Tyr Asn Ser Ile Thr Thr Ala Leu Ser Lys Val Pro Asn Ala

```
                465                 470                 475
Gln Ser Leu Gln Asn Val Val Ser Lys Lys Asn Asn Pro Tyr Ser Pro
            480                 485                 490
Gln Gly Ile Glu Thr Asn Tyr Tyr Leu Asn Gln Asn Ser Tyr Asn Gln
        495                 500                 505
Ile Gln Thr Ile Asn Gln Glu Leu Gly Arg Asn Pro Phe Arg Lys Val
    510                 515                 520
Gly Ile Val Asn Ser Gln Thr Asn Asn Gly Ala Met Asn Gly Ile Gly
525                 530                 535                 540
Ile Gln Val Gly Tyr Lys Gln Phe Phe Gly Gln Lys Arg Lys Trp Gly
                545                 550                 555
Ala Arg Tyr Tyr Gly Phe Phe Asp Tyr Asn His Ala Phe Ile Lys Ser
            560                 565                 570
Ser Phe Phe Asn Ser Ala Ser Asp Val Trp Thr Tyr Gly Phe Gly Ala
        575                 580                 585
Asp Ala Leu Tyr Asn Phe Ile Asn Asp Lys Ala Thr Asn Phe Leu Gly
    590                 595                 600
Lys Asn Asn Lys Leu Ser Leu Gly Leu Phe Gly Gly Ile Ala Leu Ala
605                 610                 615                 620
Gly Thr Ser Trp Leu Asn Ser Glu Tyr Val Asn Leu Ala Thr Val Asn
                625                 630                 635
Asn Val Tyr Asn Ala Lys Met Asn Val Ala Asn Phe Gln Phe Leu Phe
            640                 645                 650
Asn Met Gly Val Arg Met Asn Leu Ala Arg Ser Lys Lys Lys Gly Ser
        655                 660                 665
Asp His Ala Ala Gln His Gly Ile Glu Leu Gly Leu Lys Ile Pro Thr
    670                 675                 680
Ile Asn Thr Asn Tyr Tyr Ser Phe Met Gly Ala Glu Leu Lys Tyr Arg
685                 690                 695                 700
Arg Leu Tyr Ser Val Tyr Leu Asn Tyr Val Phe Ala Tyr
                705                 710
```

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2915 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 365...2597
(D) OTHER INFORMATION:

(A) NAME/KEY: Signal Sequence
(B) LOCATION: 365...425
(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
TTTTAGGCGA CAAAATCGCT TATGTTGGGG ATAAAGGCAA CCCGCACAAT TTCGCTCACA      60
AGAAATAAAC CGCTCATAAG GGGCAAACGC CCCAAAAAAG CGATTTTTAA AGAGGTTACG     120
GCAAAATCAA GCTCTTTAGT ATTTAATCTT AAAAAATGCT AAAAGCCTTT TTATGGGCTA     180
ACACCACACA AAAAGCATCA AAATCAAAAA AATGACAAAA TTTTTAAGAA AATGACAAAA     240
AAAAACGCTT TATGCTATAA TACCCCAAAT ACATTCTAAT AGCAAATGCG TTCTAATGCA     300
AATGCATTCC AATGTATGAA ATCCCTAATA CTAAATCCAA TTTAATCCAA AAAGGAGAAA     360
AAAC ATG AAA AAA CAC ATC CTT TCA TTA GCT TTA GGC TCG CTT TTA GTT       409
     Met Lys Lys His Ile Leu Ser Leu Ala Leu Gly Ser Leu Leu Val
     -20                 -15                 -10

TCC ACT TTG AGC GCT GAA GAC GAC GGC TTT TAC ACA AGC GTA GGC TAT       457
Ser Thr Leu Ser Ala Glu Asp Asp Gly Phe Tyr Thr Ser Val Gly Tyr
-5                   1               5                  10

CAG ATC GGT GAA GCC GCT CAA ATG GTA ACA AAC ACC AAA GGC ATC CAA       505
Gln Ile Gly Glu Ala Ala Gln Met Val Thr Asn Thr Lys Gly Ile Gln
            15                  20                  25

CAG CTT TCA GAC AAT TAT GAA AAT TTG AAC AAC CTT TTA ACG AGA TAC       553
Gln Leu Ser Asp Asn Tyr Glu Asn Leu Asn Asn Leu Leu Thr Arg Tyr
        30                  35                  40

AGC ACC CTA AAC ACC CTT ATC AAA TTG TCC GCT GAT CCG AGC GCA ATT       601
Ser Thr Leu Asn Thr Leu Ile Lys Leu Ser Ala Asp Pro Ser Ala Ile
    45                  50                  55

AAT GCG GTG CGG GAA AAT CTG GGC GCG AGC ACG AAG AAT TTG ATC GGC       649
Asn Ala Val Arg Glu Asn Leu Gly Ala Ser Thr Lys Asn Leu Ile Gly
60                  65                  70                  75

GAT AAA GCC AAC TCC CCG GCG TAT CAA GCC GTG TTT TTA GCG ATC AAC       697
Asp Lys Ala Asn Ser Pro Ala Tyr Gln Ala Val Phe Leu Ala Ile Asn
                80                  85                  90

GCG GCG GTA GGG TTG TGG AAT ACC ATC GGC TAT GCG GTC ATG TGC GGG       745
Ala Ala Val Gly Leu Trp Asn Thr Ile Gly Tyr Ala Val Met Cys Gly
            95                  100                 105

AAC GGG AAC GGC ACA GAG AGT GGG CCT GGC AGC GTG ATC TTT AAT GAC       793
Asn Gly Asn Gly Thr Glu Ser Gly Pro Gly Ser Val Ile Phe Asn Asp
        110                 115                 120

CAA CCA GGA CAG GAT TCC ACG CAA ATT ACT TGC AAC CGC TTT GAA TCA       841
Gln Pro Gly Gln Asp Ser Thr Gln Ile Thr Cys Asn Arg Phe Glu Ser
    125                 130                 135

ACT GGG CCT GGT AAA AGC ATG TCT ATT GAT GAA TTC AAA AAA CTC AAT       889
Thr Gly Pro Gly Lys Ser Met Ser Ile Asp Glu Phe Lys Lys Leu Asn
140                 145                 150                 155

GAA GCC TAT CAA ATC ATC CAG CAA GCT TTA AAA AAT CAA AGT GGG TTT       937
Glu Ala Tyr Gln Ile Ile Gln Gln Ala Leu Lys Asn Gln Ser Gly Phe
                160                 165                 170

CCT GAA TTA GGC GGG AAC GGC ACA AAA GTG AGT GTT AAT TAC AAT TAC       985
Pro Glu Leu Gly Gly Asn Gly Thr Lys Val Ser Val Asn Tyr Asn Tyr
            175                 180                 185
```

```
GAA TGC AGA CAA ACT GCT GAT ATC AAC GGC GGT GTG TAT CAG TTC TGC    1033
Glu Cys Arg Gln Thr Ala Asp Ile Asn Gly Gly Val Tyr Gln Phe Cys
        190                 195                 200

AAG GCT AAA AAT GGT AGT AGT AGC AGT AGT AAT GGC GGT AAT GGC AGT    1081
Lys Ala Lys Asn Gly Ser Ser Ser Ser Ser Asn Gly Gly Asn Gly Ser
    205                 210                 215

AGC ACG CAA ACA ACC GCG ACA ACC ACG CAA GAC GGC GTA ACG ATC ACC    1129
Ser Thr Gln Thr Thr Ala Thr Thr Thr Gln Asp Gly Val Thr Ile Thr
220                 225                 230                 235

ACT ACC TAT AAT AAT AAC AAA GCC ACC GTC AAA TTT GAC ATC ACC AAT    1177
Thr Thr Tyr Asn Asn Asn Lys Ala Thr Val Lys Phe Asp Ile Thr Asn
                240                 245                 250

AAC GCT GAA CAG CTG TTA AAT CAA GCG GCA AAC ATC ATG CAA GTC CTT    1225
Asn Ala Glu Gln Leu Leu Asn Gln Ala Ala Asn Ile Met Gln Val Leu
            255                 260                 265

AAT ACG CAA TGC CCT TTA GTG CGT TCC ACG AAT AAC GAA AAC ACT CCA    1273
Asn Thr Gln Cys Pro Leu Val Arg Ser Thr Asn Asn Glu Asn Thr Pro
        270                 275                 280

GGG GGT GGT CAA CCA TGG GGT TTA AGC ACA TCC GGG AAT GCG TGC AGC    1321
Gly Gly Gly Gln Pro Trp Gly Leu Ser Thr Ser Gly Asn Ala Cys Ser
    285                 290                 295

ATC TTC CAA CAA GAA TTT AGC CAG GTT ACT AGC ATG ATC AAA AAC GCC    1369
Ile Phe Gln Gln Glu Phe Ser Gln Val Thr Ser Met Ile Lys Asn Ala
300                 305                 310                 315

CAA GAA ATA ATC GCG CAA AGC AAA ATC GTT AGT GAA AAC GCG CAA AAT    1417
Gln Glu Ile Ile Ala Gln Ser Lys Ile Val Ser Glu Asn Ala Gln Asn
                320                 325                 330

CAA AAC AAC TTG GAT ACT GGA AAA CCA TTC AAC CCT TAC ACG GAC GCC    1465
Gln Asn Asn Leu Asp Thr Gly Lys Pro Phe Asn Pro Tyr Thr Asp Ala
            335                 340                 345

AGC TTT GCG CAA AGC ATG CTC AAA AAC GCT CAA GCG CAA GCA GAG ATG    1513
Ser Phe Ala Gln Ser Met Leu Lys Asn Ala Gln Ala Gln Ala Glu Met
        350                 355                 360

TTC AAT TTG AGC GAA CAA GTG AAA AAG AAC TTG GAA GTC ATG AAA AAC    1561
Phe Asn Leu Ser Glu Gln Val Lys Lys Asn Leu Glu Val Met Lys Asn
    365                 370                 375

AAC AAT AAT GTT AAC GAG AAA TTA GCA GGA TTT GGG AAA GAA GAA GTA    1609
Asn Asn Asn Val Asn Glu Lys Leu Ala Gly Phe Gly Lys Glu Glu Val
380                 385                 390                 395

ATG ACC AAT TTT GTT AGC GCC TTT TTG GCA AGC TGC AAA GAT GGT GGC    1657
Met Thr Asn Phe Val Ser Ala Phe Leu Ala Ser Cys Lys Asp Gly Gly
                400                 405                 410

ACA TTG CCT AAT GCA GGG GTT ACT TCT AAC ACT TGG GGG GCG GGT TGC    1705
```

```
Thr Leu Pro Asn Ala Gly Val Thr Ser Asn Thr Trp Gly Ala Gly Cys
            415                 420                 425

GCG TAT GTG GGA GAG ACG ATA AGC GCC CTA ACC AAC AGC ATC GCT CAC    1753
Ala Tyr Val Gly Glu Thr Ile Ser Ala Leu Thr Asn Ser Ile Ala His
        430                 435                 440

TTT GGC ACT CAA GAG CAG CAG ATA CAG CAA GCC GAA AAC ATC GCT GAC    1801
Phe Gly Thr Gln Glu Gln Gln Ile Gln Gln Ala Glu Asn Ile Ala Asp
    445                 450                 455

ACT CTA GTG AAT TTC AAA TCT AGA TAC AGC GAA TTA GGC AAC ACC TAT    1849
Thr Leu Val Asn Phe Lys Ser Arg Tyr Ser Glu Leu Gly Asn Thr Tyr
460                 465                 470                 475

AAC AGC ATC ACC ACC GCG CTC TCC AAA GTC CCT AAC GCG CAA AGC TTG    1897
Asn Ser Ile Thr Thr Ala Leu Ser Lys Val Pro Asn Ala Gln Ser Leu
                480                 485                 490

CAA AAC GTG GTG AGC AAA AAG AAT AAC CCC TAT AGC CCT CAA GGC ATA    1945
Gln Asn Val Val Ser Lys Lys Asn Asn Pro Tyr Ser Pro Gln Gly Ile
            495                 500                 505

GAG ACC AAT TAC TAC CTC AAT CAA AAT TCT TAC AAC CAA ATC CAA ACC    1993
Glu Thr Asn Tyr Tyr Leu Asn Gln Asn Ser Tyr Asn Gln Ile Gln Thr
        510                 515                 520

ATC AAC CAA GAA CTA GGG CGT AAC CCC TTT AGG AAA GTG GGC ATC GTC    2041
Ile Asn Gln Glu Leu Gly Arg Asn Pro Phe Arg Lys Val Gly Ile Val
    525                 530                 535

AAT TCT CAA ACC AAC AAT GGT GCC ATG AAT GGG ATC GGC ATT CAG GTG    2089
Asn Ser Gln Thr Asn Asn Gly Ala Met Asn Gly Ile Gly Ile Gln Val
540                 545                 550                 555

GGC TAT AAG CAA TTC TTT GGC CAA AAA AGA AAA TGG GGC GCT AGG TAT    2137
Gly Tyr Lys Gln Phe Phe Gly Gln Lys Arg Lys Trp Gly Ala Arg Tyr
                560                 565                 570

TAC GGC TTT TTT GAT TAC AAC CAT GCG TTC ATC AAA TCC AGC TTT TTC    2185
Tyr Gly Phe Phe Asp Tyr Asn His Ala Phe Ile Lys Ser Ser Phe Phe
            575                 580                 585

AAC TCG GCT TCT GAC GTG TGG ACT TAT GGT TTT GGA GCG GAC GCG CTT    2233
Asn Ser Ala Ser Asp Val Trp Thr Tyr Gly Phe Gly Ala Asp Ala Leu
        590                 595                 600

TAT AAC TTC ATC AAC GAT AAA GCC ACC AAT TTC TTA GGC AAA AAC AAC    2281
Tyr Asn Phe Ile Asn Asp Lys Ala Thr Asn Phe Leu Gly Lys Asn Asn
    605                 610                 615

AAG CTT TCT TTG GGG CTT TTT GGC GGG ATT GCG TTA GCG GGC ACT TCA    2329
Lys Leu Ser Leu Gly Leu Phe Gly Gly Ile Ala Leu Ala Gly Thr Ser
620                 625                 630                 635

TGG CTC AAT TCT GAG TAC GTG AAT TTA GCC ACC GTG AAT AAC GTC TAT    2377
Trp Leu Asn Ser Glu Tyr Val Asn Leu Ala Thr Val Asn Asn Val Tyr
```

```
                640                 645                 650

AAC GCT AAA ATG AAT GTG GCG AAT TTC CAA TTC TTA TTC AAT ATG GGA    2425
Asn Ala Lys Met Asn Val Ala Asn Phe Gln Phe Leu Phe Asn Met Gly
            655                 660                 665

GTG AGG ATG AAT TTA GCC AGA TCC AAG AAA AAA GGC AGC GAT CAT GCA    2473
Val Arg Met Asn Leu Ala Arg Ser Lys Lys Lys Gly Ser Asp His Ala
        670                 675                 680

GCT CAG CAT GGG ATT GAG TTA GGG CTT AAA ATC CCC ACC ATC AAC ACG    2521
Ala Gln His Gly Ile Glu Leu Gly Leu Lys Ile Pro Thr Ile Asn Thr
    685                 690                 695

AAC TAT TAT TCC TTT ATG GGG GCT GAA CTC AAA TAC AGA AGG CTC TAT    2569
Asn Tyr Tyr Ser Phe Met Gly Ala Glu Leu Lys Tyr Arg Arg Leu Tyr
700                 705                 710                 715

AGC GTG TAT TTG AAT NAT GTG TTC GCT TAC TAAGCTTTTT GTGAAACTCC     2619
Ser Val Tyr Leu Asn Xaa Val Phe Ala Tyr
                720                 725

CTTTTTAAGG GGTTTTTTTT TGAACTCTCT TTTAAATTCT CTTTTTAAAG AGATTTCTTT  2679
TTTTAAGCTT TTTTTTGAAC TTTTTTTTGA ATTCTTTGTT TTTAAGCTTT TTTTAAACCC  2739
TTTCGTTTTT AAACTCCCTT TTTTAAGGGA TTTCTTTTTT TGAACTCCCT TTTTTGAACC  2799
CTTTTTTTTA AACCCTCTTT TTTTAAGGGG TTTCTTTTTA AAGCTTTTTT GAAGTCTTTT  2859
TTTAAATTCT TTTTTTGGGG GTTTGATCTT TCTTTTTGCC AATCCCCACT ACTTTC      2915
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 745 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal
(ix) FEATURE:

(A) NAME/KEY: Signal Sequence
(B) LOCATION: 1...20
(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
Met Lys Lys His Ile Leu Ser Leu Ala Leu Gly Ser Leu Leu Val Ser
-20                 -15                 -10                 -5
Thr Leu Ser Ala Glu Asp Asp Gly Phe Tyr Thr Ser Val Gly Tyr Gln
                 1               5                  10
Ile Gly Glu Ala Ala Gln Met Val Thr Asn Thr Lys Gly Ile Gln Gln
        15                  20                  25
Leu Ser Asp Asn Tyr Glu Asn Leu Asn Asn Leu Leu Thr Arg Tyr Ser
```

```
    30                  35                  40
Thr Leu Asn Thr Leu Ile Lys Leu Ser Ala Asp Pro Ser Ala Ile Asn
45                  50                  55                  60
Ala Val Arg Glu Asn Leu Gly Ala Ser Thr Lys Asn Leu Ile Gly Asp
                65                  70                  75
Lys Ala Asn Ser Pro Ala Tyr Gln Ala Val Phe Leu Ala Ile Asn Ala
            80                  85                  90
Ala Val Gly Leu Trp Asn Thr Ile Gly Tyr Ala Val Met Cys Gly Asn
        95                  100                 105
Gly Asn Gly Thr Glu Ser Gly Pro Gly Ser Val Ile Phe Asn Asp Gln
    110                 115                 120
Pro Gly Gln Asp Ser Thr Gln Ile Thr Cys Asn Arg Phe Glu Ser Thr
125                 130                 135                 140
Gly Pro Gly Lys Ser Met Ser Ile Asp Glu Phe Lys Lys Leu Asn Glu
                145                 150                 155
Ala Tyr Gln Ile Ile Gln Gln Ala Leu Lys Asn Gln Ser Gly Phe Pro
            160                 165                 170
Glu Leu Gly Gly Asn Gly Thr Lys Val Ser Val Asn Tyr Asn Tyr Glu
        175                 180                 185
Cys Arg Gln Thr Ala Asp Ile Asn Gly Gly Val Tyr Gln Phe Cys Lys
    190                 195                 200
Ala Lys Asn Gly Ser Ser Ser Ser Ser Asn Gly Gly Asn Gly Ser Ser
205                 210                 215                 220
Thr Gln Thr Thr Ala Thr Thr Thr Gln Asp Gly Val Thr Ile Thr Thr
                225                 230                 235
Thr Tyr Asn Asn Asn Lys Ala Thr Val Lys Phe Asp Ile Thr Asn Asn
            240                 245                 250
Ala Glu Gln Leu Leu Asn Gln Ala Ala Asn Ile Met Gln Val Leu Asn
        255                 260                 265
Thr Gln Cys Pro Leu Val Arg Ser Thr Asn Asn Glu Asn Thr Pro Gly
    270                 275                 280
Gly Gly Gln Pro Trp Gly Leu Ser Thr Ser Gly Asn Ala Cys Ser Ile
285                 290                 295                 300
Phe Gln Gln Glu Phe Ser Gln Val Thr Ser Met Ile Lys Asn Ala Gln
                305                 310                 315
Glu Ile Ile Ala Gln Ser Lys Ile Val Ser Glu Asn Ala Gln Asn Gln
            320                 325                 330
Asn Asn Leu Asp Thr Gly Lys Pro Phe Asn Pro Tyr Thr Asp Ala Ser
        335                 340                 345
Phe Ala Gln Ser Met Leu Lys Asn Ala Gln Ala Gln Ala Glu Met Phe
    350                 355                 360
Asn Leu Ser Glu Gln Val Lys Lys Asn Leu Glu Val Met Lys Asn Asn
365                 370                 375                 380
Asn Asn Val Asn Glu Lys Leu Ala Gly Phe Gly Lys Glu Glu Val Met
                385                 390                 395
Thr Asn Phe Val Ser Ala Phe Leu Ala Ser Cys Lys Asp Gly Gly Thr
            400                 405                 410
Leu Pro Asn Ala Gly Val Thr Ser Asn Thr Trp Gly Ala Gly Cys Ala
        415                 420                 425
Tyr Val Gly Glu Thr Ile Ser Ala Leu Thr Asn Ser Ile Ala His Phe
    430                 435                 440
Gly Thr Gln Glu Gln Gln Ile Gln Gln Ala Glu Asn Ile Ala Asp Thr
445                 450                 455                 460
Leu Val Asn Phe Lys Ser Arg Tyr Ser Glu Leu Gly Asn Thr Tyr Asn
                465                 470                 475
Ser Ile Thr Thr Ala Leu Ser Lys Val Pro Asn Ala Gln Ser Leu Gln
            480                 485                 490
```

```
Asn Val Val Ser Lys Lys Asn Asn Pro Tyr Ser Pro Gln Gly Ile Glu
        495                 500                 505
Thr Asn Tyr Tyr Leu Asn Gln Asn Ser Tyr Asn Gln Ile Gln Thr Ile
    510                 515                 520
Asn Gln Glu Leu Gly Arg Asn Pro Phe Arg Lys Val Gly Ile Val Asn
525                 530                 535                 540
Ser Gln Thr Asn Asn Gly Ala Met Asn Gly Ile Gly Ile Gln Val Gly
                545                 550                 555
Tyr Lys Gln Phe Phe Gly Gln Lys Arg Lys Trp Gly Ala Arg Tyr Tyr
            560                 565                 570
Gly Phe Phe Asp Tyr Asn His Ala Phe Ile Lys Ser Ser Phe Phe Asn
        575                 580                 585
Ser Ala Ser Asp Val Trp Thr Tyr Gly Phe Gly Ala Asp Ala Leu Tyr
    590                 595                 600
Asn Phe Ile Asn Asp Lys Ala Thr Asn Phe Leu Gly Lys Asn Asn Lys
605                 610                 615                 620
Leu Ser Leu Gly Leu Phe Gly Gly Ile Ala Leu Ala Gly Thr Ser Trp
                625                 630                 635
Leu Asn Ser Glu Tyr Val Asn Leu Ala Thr Val Asn Asn Val Tyr Asn
            640                 645                 650
Ala Lys Met Asn Val Ala Asn Phe Gln Phe Leu Phe Asn Met Gly Val
        655                 660                 665
Arg Met Asn Leu Ala Arg Ser Lys Lys Lys Gly Ser Asp His Ala Ala
    670                 675                 680
Gln His Gly Ile Glu Leu Gly Leu Lys Ile Pro Thr Ile Asn Thr Asn
685                 690                 695                 700
Tyr Tyr Ser Phe Met Gly Ala Glu Leu Lys Tyr Arg Arg Leu Tyr Ser
                705                 710                 715
Val Tyr Leu Asn Xaa Val Phe Ala Tyr
            720                 725
```

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2603 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 210...2342
(D) OTHER INFORMATION:

(A) NAME/KEY: Signal Sequence
(B) LOCATION: 210...270
(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
ATGACCTTTA TTGGTTTAAT ATTTGTTTAG AAATAACACA AAAACCTTTT TTTTTTTTTT      60
```

```
TGAAAGGGCA AAAACGCCTA ATTAATATCA AAATCCCATG AATTTATACT ATATTAACGA   120
AAGCTTGCGG TATGGTTTCA CCTAAAGACA CACTTCCGCA AGATTTACTA ACAATTTCAA   180
TCTTATTTCA AGTAATAAAA GGAGAAAAC ATG AAG AAA AAA TTT CTG TCA TTA      233
                                Met Lys Lys Lys Phe Leu Ser Leu
                                -20                 -15

ACC TTA GGT TCG CTT TTA GTT TCC GCT TTA AGC GCT GAA GAC AAC GGC      281
Thr Leu Gly Ser Leu Leu Val Ser Ala Leu Ser Ala Glu Asp Asn Gly
        -10                 -5                  1

TTT TTT GTG AGT GCG GGC TAT CAA ATC GGT GAA TCC GCT CAA ATG GTG      329
Phe Phe Val Ser Ala Gly Tyr Gln Ile Gly Glu Ser Ala Gln Met Val
 5                  10                  15                  20

AAA AAC ACT AAA GGC ATT CAA GAT CTT TCA GAT AGC TAT GAA AGA CTG      377
Lys Asn Thr Lys Gly Ile Gln Asp Leu Ser Asp Ser Tyr Glu Arg Leu
                25                  30                  35

AAC AAT CTT TTA ACG AGT TAT AGT GCC CTA AAC ACT CTT ATT AGG CAG      425
Asn Asn Leu Leu Thr Ser Tyr Ser Ala Leu Asn Thr Leu Ile Arg Gln
            40                  45                  50

TCC GCC GAC CCC AAC GCT ATC AAT AAC GCA AGG GGC AAT TTG AAC GCT      473
Ser Ala Asp Pro Asn Ala Ile Asn Asn Ala Arg Gly Asn Leu Asn Ala
        55                  60                  65

AGT GCG AAG AAT TTG ATC AAT GAT AAA AAG AAT TCC CCG GCG TAT CAA      521
Ser Ala Lys Asn Leu Ile Asn Asp Lys Lys Asn Ser Pro Ala Tyr Gln
    70                  75                  80

GCG GTG CTT TTA GCC TTG AAT GCG GCA GCG GGG TTG TGG CAA GTC ATG      569
Ala Val Leu Leu Ala Leu Asn Ala Ala Ala Gly Leu Trp Gln Val Met
85                  90                  95                  100

AGC TAT TCG ATC AGC GTT TGT GGC CCT GGC TCT GAC AAA AAT AAA AAT      617
Ser Tyr Ser Ile Ser Val Cys Gly Pro Gly Ser Asp Lys Asn Lys Asn
                105                 110                 115

GGG GGC GTC CAA ACC TTT GAA AAT GTG CCG TCA AAT GGG GGG ACT ACC      665
Gly Gly Val Gln Thr Phe Glu Asn Val Pro Ser Asn Gly Gly Thr Thr
            120                 125                 130

ATT GCT TGC GAT TCA TTT TAT GAA CCA GGA AAG TGG AGC GGT ATA TCC      713
Ile Ala Cys Asp Ser Phe Tyr Glu Pro Gly Lys Trp Ser Gly Ile Ser
        135                 140                 145

ACT GAA AAT TAC GCA AAA ATC AAT AAA GCC TAT CAA ATC ATC CAA AAG      761
Thr Glu Asn Tyr Ala Lys Ile Asn Lys Ala Tyr Gln Ile Ile Gln Lys
    150                 155                 160

GCT TTT GGA GCA AGC GGG CAA GAT ATT CCT GCC TTA AGC GAC ACC AAA      809
Ala Phe Gly Ala Ser Gly Gln Asp Ile Pro Ala Leu Ser Asp Thr Lys
165                 170                 175                 180

GAA CTT AAT TTT GAA ATT AAA GGG AAA AAA AAT GAT AGC GTC CAG CCA      857
Glu Leu Asn Phe Glu Ile Lys Gly Lys Lys Asn Asp Ser Val Gln Pro
                185                 190                 195
```

```
GGA GAA AGA TGG AAA TTC CCA TGG ACT AAT GGA AAA TTT GTT TCA GTC      905
Gly Glu Arg Trp Lys Phe Pro Trp Thr Asn Gly Lys Phe Val Ser Val
            200                 205                 210

AAG TGG GTG AAT GGG AAG TAT GAA GAA ATT AAA GAA GAC ATC AAA GTG      953
Lys Trp Val Asn Gly Lys Tyr Glu Glu Ile Lys Glu Asp Ile Lys Val
        215                 220                 225

TCA AAT AAC GCT CAA GAG CTT TTA AAA CAG GCT AGC ACT ATT TTA ACC     1001
Ser Asn Asn Ala Gln Glu Leu Leu Lys Gln Ala Ser Thr Ile Leu Thr
    230                 235                 240

ACT CTT AAT GAA GCA TGC CCA TGG TTG AGT AAT GGT GGT GCA GGC AAT     1049
Thr Leu Asn Glu Ala Cys Pro Trp Leu Ser Asn Gly Gly Ala Gly Asn
245                 250                 255                 260

GTG GCC GGT GGC AAT AGT TTA TGG GCC GGA ATA GAT AAA GGC GAC GGG     1097
Val Ala Gly Gly Asn Ser Leu Trp Ala Gly Ile Asp Lys Gly Asp Gly
                265                 270                 275

AGC GCA TGC GGG ATT TTT AAA AAT GAA ATC AGC GCG ATT CAA GAC ATG     1145
Ser Ala Cys Gly Ile Phe Lys Asn Glu Ile Ser Ala Ile Gln Asp Met
            280                 285                 290

ATC AAA AAC GCT GAA ATA GCC GTA GAG CAA TCC AAA ATC GTT ACC GCC     1193
Ile Lys Asn Ala Glu Ile Ala Val Glu Gln Ser Lys Ile Val Thr Ala
        295                 300                 305

AAC GCG CAA AAC CAG CAC AAC CTA GAC ACT GGG AAA GCA TTC AAC CCC     1241
Asn Ala Gln Asn Gln His Asn Leu Asp Thr Gly Lys Ala Phe Asn Pro
    310                 315                 320

TAT AAA GAC GCC AAC TTC GCC CAA AGC ATG TTC GCT AAC GCT AGA GCG     1289
Tyr Lys Asp Ala Asn Phe Ala Gln Ser Met Phe Ala Asn Ala Arg Ala
325                 330                 335                 340

CAA GCG GAG ATT TTA AAC CGC GCT CAA GCA GTG GTG AAG GAC TTT GAA     1337
Gln Ala Glu Ile Leu Asn Arg Ala Gln Ala Val Val Lys Asp Phe Glu
                345                 350                 355

AGA ATC CCT GCA GCG TTC GTG AAA GAC TCT TTA GGA GTA TGC CAT GAA     1385
Arg Ile Pro Ala Ala Phe Val Lys Asp Ser Leu Gly Val Cys His Glu
            360                 365                 370

AAG GGT AGC GAC GGC AAT CTC CGT GGC ACG CCA TCT GGC ACG GTT ACT     1433
Lys Gly Ser Asp Gly Asn Leu Arg Gly Thr Pro Ser Gly Thr Val Thr
        375                 380                 385

TCT AAC ACT TGG GGA GCC GGC TGC GCG TAT GTG GGA GAA ACC GTA ACG     1481
Ser Asn Thr Trp Gly Ala Gly Cys Ala Tyr Val Gly Glu Thr Val Thr
    390                 395                 400

AAT CTA AAA AAC AGC ATC GCT CAT TTT GGC GAC CAA GCG GAG CGA ATC     1529
Asn Leu Lys Asn Ser Ile Ala His Phe Gly Asp Gln Ala Glu Arg Ile
405                 410                 415                 420

CAT AAT GCG CGA AAT CTC GCC TAC ACT TTA GCG AAT TTC AGC GGC CAG     1577
```

```
His Asn Ala Arg Asn Leu Ala Tyr Thr Leu Ala Asn Phe Ser Gly Gln
                425                 430                 435

TAC AAA AAG CTA GGC GAA CAC TAT GAC AGC ATC ACA GCG GCG CTC TCT    1625
Tyr Lys Lys Leu Gly Glu His Tyr Asp Ser Ile Thr Ala Ala Leu Ser
            440                 445                 450

AGC TTG CCT GAT GCG CAA TCT TTA CAA AAT GTG GTG AGC AAA AAG ACT    1673
Ser Leu Pro Asp Ala Gln Ser Leu Gln Asn Val Val Ser Lys Lys Thr
        455                 460                 465

AAC CCT AAC AGC CCG CAA GGC ATA CAG GAT AAT TAC TAC ATT GAC TCC    1721
Asn Pro Asn Ser Pro Gln Gly Ile Gln Asp Asn Tyr Tyr Ile Asp Ser
    470                 475                 480

AAC ATC CAT TCT CAA GTG CAA TCT AGG AGT CAA GAA CTC GGC AGT AAC    1769
Asn Ile His Ser Gln Val Gln Ser Arg Ser Gln Glu Leu Gly Ser Asn
485                 490                 495                 500

CCT TTC AGA CGC GCC GGG CTA ATC GCC GCT TCT ACC ACC AAT AAC GGC    1817
Pro Phe Arg Arg Ala Gly Leu Ile Ala Ala Ser Thr Thr Asn Asn Gly
                505                 510                 515

GCG ATG AAT GGG ATT GGC TTT CAA GTG GGC TAT AAG CAA TTC TTT GGG    1865
Ala Met Asn Gly Ile Gly Phe Gln Val Gly Tyr Lys Gln Phe Phe Gly
            520                 525                 530

AAA AAC AAA CGA TGG GGC GCG AGA TAC TAC GGC TTT GTG GAT TAC AAC    1913
Lys Asn Lys Arg Trp Gly Ala Arg Tyr Tyr Gly Phe Val Asp Tyr Asn
        535                 540                 545

CAC ACC TAT AAC AAG TCC CAA TTT TTC AAC TCC GAT TCT GAT GTT TGG    1961
His Thr Tyr Asn Lys Ser Gln Phe Phe Asn Ser Asp Ser Asp Val Trp
    550                 555                 560

ACT TAT GGC GTG GGG AGC GAT TTG TTA GTG AAT TTC ATC AAC GAT AAA    2009
Thr Tyr Gly Val Gly Ser Asp Leu Leu Val Asn Phe Ile Asn Asp Lys
565                 570                 575                 580

GCC ACT AAA CAC AAT AAA ATT TCT TTT GGC GCG TTT GGC GGT ATC CAA    2057
Ala Thr Lys His Asn Lys Ile Ser Phe Gly Ala Phe Gly Gly Ile Gln
                585                 590                 595

CTA GCC GGG ACT TCA TGG CTT AAT TCT CAG TAT GTG AAT TTA GCG AAT    2105
Leu Ala Gly Thr Ser Trp Leu Asn Ser Gln Tyr Val Asn Leu Ala Asn
            600                 605                 610

GTG AAC AAT TAT TAT AAA GCT AAA ATC AAC ACC TCT AAC TTC CAA TTC    2153
Val Asn Asn Tyr Tyr Lys Ala Lys Ile Asn Thr Ser Asn Phe Gln Phe
        615                 620                 625

TTA TTC AAT CTG GGC TTA AGG ACC AAT CTC GCC AGA AAT AAA AGA ATA    2201
Leu Phe Asn Leu Gly Leu Arg Thr Asn Leu Ala Arg Asn Lys Arg Ile
    630                 635                 640

GGC GCT GAT CAT AGC GCG CAA CAT GGC ATG GAA TTA GGC GTG AAG ATC    2249
Gly Ala Asp His Ser Ala Gln His Gly Met Glu Leu Gly Val Lys Ile
```

```
645                 650                 655                 660

CCC ACG ATC AAC ACA AAT TAC TAT TCT TTG CTA GGC ACT ACC TTG CAA    2297
Pro Thr Ile Asn Thr Asn Tyr Tyr Ser Leu Leu Gly Thr Thr Leu Gln
                665                 670                 675

TAC AGA AGG CTT TAT AGC GTG TAT CTC AAC TAT GTG TTT GCT TAC TAAAA  2347
Tyr Arg Arg Leu Tyr Ser Val Tyr Leu Asn Tyr Val Phe Ala Tyr
            680                 685                 690

GCTTAAACTC CTTTTTAAAC TCCCTTTTTA GGGGGTTTAA TCTTTTTAAC TGACTTTTCT   2407
TTTAGCTTTT TTTAATTTTT TCCACCAAAC AAAGTTTTTT GACTTCAAGC GTTAATCACA   2467
AAAAATACTC AAAGGCGTTT TTTGCAATCT AAATAAAAAA TTAGCGTTAT TCAAGCGATC   2527
ATTTTAAACC ACCCAAGCAA GAAACCCCAA ACATCTTTAG CGTTCGCGCG CTCCACTAAC   2587
CAAAAAACGC CCCAAA                                                   2603
```

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 711 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal
(ix) FEATURE:

(A) NAME/KEY: Signal Sequence
(B) LOCATION: 1...20
(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```
Met Lys Lys Lys Phe Leu Ser Leu Thr Leu Gly Ser Leu Leu Val Ser
-20                 -15                 -10                 -5
Ala Leu Ser Ala Glu Asp Asn Gly Phe Phe Val Ser Ala Gly Tyr Gln
                 1                   5                  10
Ile Gly Glu Ser Ala Gln Met Val Lys Asn Thr Lys Gly Ile Gln Asp
        15                  20                  25
Leu Ser Asp Ser Tyr Glu Arg Leu Asn Asn Leu Leu Thr Ser Tyr Ser
    30                  35                  40
Ala Leu Asn Thr Leu Ile Arg Gln Ser Ala Asp Pro Asn Ala Ile Asn
45                  50                  55                  60
Asn Ala Arg Gly Asn Leu Asn Ala Ser Ala Lys Asn Leu Ile Asn Asp
                65                  70                  75
Lys Lys Asn Ser Pro Ala Tyr Gln Ala Val Leu Leu Ala Leu Asn Ala
            80                  85                  90
Ala Ala Gly Leu Trp Gln Val Met Ser Tyr Ser Ile Ser Val Cys Gly
        95                  100                 105
Pro Gly Ser Asp Lys Asn Lys Asn Gly Gly Val Gln Thr Phe Glu Asn
    110                 115                 120
Val Pro Ser Asn Gly Gly Thr Thr Ile Ala Cys Asp Ser Phe Tyr Glu
125                 130                 135                 140
Pro Gly Lys Trp Ser Gly Ile Ser Thr Glu Asn Tyr Ala Lys Ile Asn
```

```
                145                 150                 155
Lys Ala Tyr Gln Ile Ile Gln Lys Ala Phe Gly Ala Ser Gly Gln Asp
            160                 165                 170
Ile Pro Ala Leu Ser Asp Thr Lys Glu Leu Asn Phe Glu Ile Lys Gly
        175                 180                 185
Lys Lys Asn Asp Ser Val Gln Pro Gly Glu Arg Trp Lys Phe Pro Trp
    190                 195                 200
Thr Asn Gly Lys Phe Val Ser Val Lys Trp Val Asn Gly Lys Tyr Glu
205                 210                 215                 220
Glu Ile Lys Glu Asp Ile Lys Val Ser Asn Asn Ala Gln Glu Leu Leu
                225                 230                 235
Lys Gln Ala Ser Thr Ile Leu Thr Thr Leu Asn Glu Ala Cys Pro Trp
            240                 245                 250
Leu Ser Asn Gly Gly Ala Gly Asn Val Ala Gly Gly Asn Ser Leu Trp
        255                 260                 265
Ala Gly Ile Asp Lys Gly Asp Gly Ser Ala Cys Gly Ile Phe Lys Asn
    270                 275                 280
Glu Ile Ser Ala Ile Gln Asp Met Ile Lys Asn Ala Glu Ile Ala Val
285                 290                 295                 300
Glu Gln Ser Lys Ile Val Thr Ala Asn Ala Gln Asn Gln His Asn Leu
                305                 310                 315
Asp Thr Gly Lys Ala Phe Asn Pro Tyr Lys Asp Ala Asn Phe Ala Gln
            320                 325                 330
Ser Met Phe Ala Asn Ala Arg Ala Gln Ala Glu Ile Leu Asn Arg Ala
        335                 340                 345
Gln Ala Val Val Lys Asp Phe Glu Arg Ile Pro Ala Ala Phe Val Lys
    350                 355                 360
Asp Ser Leu Gly Val Cys His Glu Lys Gly Ser Asp Gly Asn Leu Arg
365                 370                 375                 380
Gly Thr Pro Ser Gly Thr Val Thr Ser Asn Thr Trp Gly Ala Gly Cys
                385                 390                 395
Ala Tyr Val Gly Glu Thr Val Thr Asn Leu Lys Asn Ser Ile Ala His
            400                 405                 410
Phe Gly Asp Gln Ala Glu Arg Ile His Asn Ala Arg Asn Leu Ala Tyr
        415                 420                 425
Thr Leu Ala Asn Phe Ser Gly Gln Tyr Lys Lys Leu Gly Glu His Tyr
    430                 435                 440
Asp Ser Ile Thr Ala Ala Leu Ser Ser Leu Pro Asp Ala Gln Ser Leu
445                 450                 455                 460
Gln Asn Val Val Ser Lys Lys Thr Asn Pro Asn Ser Pro Gln Gly Ile
                465                 470                 475
Gln Asp Asn Tyr Tyr Ile Asp Ser Asn Ile His Ser Gln Val Gln Ser
            480                 485                 490
Arg Ser Gln Glu Leu Gly Ser Asn Pro Phe Arg Arg Ala Gly Leu Ile
        495                 500                 505
Ala Ala Ser Thr Thr Asn Asn Gly Ala Met Asn Gly Ile Gly Phe Gln
    510                 515                 520
Val Gly Tyr Lys Gln Phe Phe Gly Lys Asn Lys Arg Trp Gly Ala Arg
525                 530                 535                 540
Tyr Tyr Gly Phe Val Asp Tyr Asn His Thr Tyr Asn Lys Ser Gln Phe
                545                 550                 555
Phe Asn Ser Asp Ser Asp Val Trp Thr Tyr Gly Val Gly Ser Asp Leu
            560                 565                 570
Leu Val Asn Phe Ile Asn Asp Lys Ala Thr Lys His Asn Lys Ile Ser
        575                 580                 585
Phe Gly Ala Phe Gly Gly Ile Gln Leu Ala Gly Thr Ser Trp Leu Asn
    590                 595                 600
```

```
Ser Gln Tyr Val Asn Leu Ala Asn Val Asn Asn Tyr Tyr Lys Ala Lys
605                 610                 615                 620
Ile Asn Thr Ser Asn Phe Gln Phe Leu Phe Asn Leu Gly Leu Arg Thr
                625                 630                 635
Asn Leu Ala Arg Asn Lys Arg Ile Gly Ala Asp His Ser Ala Gln His
            640                 645                 650
Gly Met Glu Leu Gly Val Lys Ile Pro Thr Ile Asn Thr Asn Tyr Tyr
        655                 660                 665
Ser Leu Leu Gly Thr Thr Leu Gln Tyr Arg Arg Leu Tyr Ser Val Tyr
    670                 675                 680
Leu Asn Tyr Val Phe Ala Tyr
685                 690
```

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2427 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 232...2247
(D) OTHER INFORMATION:

(A) NAME/KEY: Signal Sequence
(B) LOCATION: 232...292
(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

```
AAAACGCGCA GCAAAAAATC TCTGTTAAGC TTTTATCATT AGCGTTCCAT TGAAACAAAA     60
TCTAAAAACC CTTTCCAATA CCACCCAAAC AAACGCGCAA AAAATGCAAA AATTCTAAAT    120
TTTCTCCAAA TGACAAAAAA AAAAAAAACG ATTTTATGCT ACAATGCTTT TAATACATTC    180
TTACTTAATG TATAAAATCT CAATCACTCA ATTTAATTTC AAAGGATATT T ATG AAA     237
                                                          Met Lys
                                                          -20

AAA ACC CTT TTA CTC TCT CTC TCT CTC TCT CTC TCG TCA TCG CTT TTA      285
Lys Thr Leu Leu Leu Ser Leu Ser Leu Ser Leu Ser Ser Ser Leu Leu
            -15                 -10                 -5

AAC GCT GAA GAC AAC GGC TTT TTT ATC AGC GCG GGC TAT CAA ATC GGT      333
Asn Ala Glu Asp Asn Gly Phe Phe Ile Ser Ala Gly Tyr Gln Ile Gly
         1               5                  10

GAA GCC GCT CAA ATG GTG AAA AAC ACC GGC GAA TTG AAA AAA CTT TCA      381
Glu Ala Ala Gln Met Val Lys Asn Thr Gly Glu Leu Lys Lys Leu Ser
15                  20                  25                  30
```

```
GAC ACT TAT GAG AAT TTG AGC AAC CTT TTA ACC AAT TTT AAC AAC CTC      429
Asp Thr Tyr Glu Asn Leu Ser Asn Leu Leu Thr Asn Phe Asn Asn Leu
                35                  40                  45

AAT CAA GCG GTA ACG AAC GCG AGC AGC CCT TCA GAA ATC AAT GCC ACG      477
Asn Gln Ala Val Thr Asn Ala Ser Ser Pro Ser Glu Ile Asn Ala Thr
            50                  55                  60

ATC GAT AAT TTA AAA GCA AAC ACG CAA GGG CTG ATT GGC GAA AAA ACC      525
Ile Asp Asn Leu Lys Ala Asn Thr Gln Gly Leu Ile Gly Glu Lys Thr
        65                  70                  75

AAT TCC CCG GCG TAT CAA GCG GTG TAT TTG GCG CTC AAT GCG GCG GTG      573
Asn Ser Pro Ala Tyr Gln Ala Val Tyr Leu Ala Leu Asn Ala Ala Val
    80                  85                  90

GGG CTG TGG AAT GTG ATA GCC TAT AAT GTC CAA TGC GGT CCT GGT AAG      621
Gly Leu Trp Asn Val Ile Ala Tyr Asn Val Gln Cys Gly Pro Gly Lys
95                  100                 105                 110

AGT GGG GAT CAA AGC GTA ATT TTT GAT GGC CAA CCA GGA CAT GAT TCA      669
Ser Gly Asp Gln Ser Val Ile Phe Asp Gly Gln Pro Gly His Asp Ser
                115                 120                 125

AGA TCC ATT AAT TGC AAT TTA ACC GGT TAT AAC AAC GGG GTT AGC GGC      717
Arg Ser Ile Asn Cys Asn Leu Thr Gly Tyr Asn Asn Gly Val Ser Gly
            130                 135                 140

CCT TTA TCC ATT GAC AAT TTT AAA ACG CTT AAT CAA GCT TAT CAA ACT      765
Pro Leu Ser Ile Asp Asn Phe Lys Thr Leu Asn Gln Ala Tyr Gln Thr
        145                 150                 155

ATC CAA CAA GCT TTA AAA CAA GAT AGC GGA TTT CCT GTT TTG GAT AGT      813
Ile Gln Gln Ala Leu Lys Gln Asp Ser Gly Phe Pro Val Leu Asp Ser
    160                 165                 170

AAA GGA AAA CAA GTA ACT ATA AAA ATA ACA ACA CAA ACT AAT GGA GCT      861
Lys Gly Lys Gln Val Thr Ile Lys Ile Thr Thr Gln Thr Asn Gly Ala
175                 180                 185                 190

AAT AAA AGT GAA ACT ACT ACT ACT ACT ACT ACT ACT AAT GAC GCT CAA      909
Asn Lys Ser Glu Thr Thr Thr Thr Thr Thr Thr Thr Asn Asp Ala Gln
                195                 200                 205

ACC CTT TTG CAA GAA GCC AGT AAA ATG ATA AGC GTC CTC ACT ACA AAC      957
Thr Leu Leu Gln Glu Ala Ser Lys Met Ile Ser Val Leu Thr Thr Asn
            210                 215                 220

TGC CCA TGG GTA AAT ACC GCT CAT AAC TCA AAC GGG GGT GCA CCG TGG     1005
Cys Pro Trp Val Asn Thr Ala His Asn Ser Asn Gly Gly Ala Pro Trp
        225                 230                 235

AAT TTA AAT ACG ACA GGG AAT GTG TGT CAG GTT TTT GCC ACG GAG TTT     1053
Asn Leu Asn Thr Thr Gly Asn Val Cys Gln Val Phe Ala Thr Glu Phe
    240                 245                 250

AGC GCC GTT ACT AGC ATG ATC AAA AAC GCG CAA GAA ATC GTA ACG CAA     1101
```

```
Ser Ala Val Thr Ser Met Ile Lys Asn Ala Gln Glu Ile Val Thr Gln
255                 260                 265                 270

GCT CAA AGC CTT AAC AAC CCG CAA AGC AAT CAA AAC GCG CCG AAA GAT    1149
Ala Gln Ser Leu Asn Asn Pro Gln Ser Asn Gln Asn Ala Pro Lys Asp
                275                 280                 285

TTC AAT CCT TAC ACC TCT GCT GAT AGG GCT TTC GCT CAA AAC ATG CTC    1197
Phe Asn Pro Tyr Thr Ser Ala Asp Arg Ala Phe Ala Gln Asn Met Leu
            290                 295                 300

AAT CAC GCG CAA GCG CAA GCC AAG ATG CTT GAA CTA GCC GAT CAA ATG    1245
Asn His Ala Gln Ala Gln Ala Lys Met Leu Glu Leu Ala Asp Gln Met
        305                 310                 315

AAA AAA GAC CTT AAC ACT ATC CCA AAA CAA TTT ATC ACA AAC TAC TTG    1293
Lys Lys Asp Leu Asn Thr Ile Pro Lys Gln Phe Ile Thr Asn Tyr Leu
    320                 325                 330

GCA GCT TGC CGC AAT GGG GGT GGG ACA TTA CCT GAT GCA GGG GTT ACT    1341
Ala Ala Cys Arg Asn Gly Gly Gly Thr Leu Pro Asp Ala Gly Val Thr
335                 340                 345                 350

TCT AAC ACT TGG GGG GCC GGT TGC GCC TAT GTG GAA GAG ACG ATA ACC    1389
Ser Asn Thr Trp Gly Ala Gly Cys Ala Tyr Val Glu Glu Thr Ile Thr
                355                 360                 365

GCC CTA AAT AAC AGC CTT GCG CAT TTT GGC ACT CAA GCC GAT CAA ATC    1437
Ala Leu Asn Asn Ser Leu Ala His Phe Gly Thr Gln Ala Asp Gln Ile
            370                 375                 380

AAG CAA TCT GAG TTG TTG GCG CGC ACG ATA CTT GAT TTT AGA GGC AGC    1485
Lys Gln Ser Glu Leu Leu Ala Arg Thr Ile Leu Asp Phe Arg Gly Ser
        385                 390                 395

CTT AAG GAT TTA AAC AAC ACT TAT AAC AGC ATC ACC ACG ACC GCT TCA    1533
Leu Lys Asp Leu Asn Asn Thr Tyr Asn Ser Ile Thr Thr Thr Ala Ser
    400                 405                 410

AAC ACG CCC AAT TCC CCA TTC CTT AAA AAT TTG ATA AGC CAA TCC ACT    1581
Asn Thr Pro Asn Ser Pro Phe Leu Lys Asn Leu Ile Ser Gln Ser Thr
415                 420                 425                 430

AAC CCT AAT AAC CCC GGG GGC TTA CAG GCC GTT TAT CAA GTC AAC CAA    1629
Asn Pro Asn Asn Pro Gly Gly Leu Gln Ala Val Tyr Gln Val Asn Gln
                435                 440                 445

AGC GCT TAT TCG CAA TTA TTA AGC GCC ACG CAA GAA TTA GGG CAT AAC    1677
Ser Ala Tyr Ser Gln Leu Leu Ser Ala Thr Gln Glu Leu Gly His Asn
            450                 455                 460

CCT TTC AGA CGC GTT GGC TTA ATC AGC TCT CAA ACC AAC AAC GGT GCG    1725
Pro Phe Arg Arg Val Gly Leu Ile Ser Ser Gln Thr Asn Asn Gly Ala
        465                 470                 475

ATG AAT GGG ATC GGC GTG CAA ATA GGG TAT AAA CAA TTT TTT GGT GAA    1773
Met Asn Gly Ile Gly Val Gln Ile Gly Tyr Lys Gln Phe Phe Gly Glu
```

```
    480                 485                 490

AAA AGA AGA TGG GGG TTA AGG TAT TAT GGT TTT TTT GAT TAC AAC CAT    1821
Lys Arg Arg Trp Gly Leu Arg Tyr Tyr Gly Phe Phe Asp Tyr Asn His
495                 500                 505                 510

GCT TAT ATC AAA TCC AGC TTT TTC AAC TCC GCC TCT GAT GTG TTC ACT    1869
Ala Tyr Ile Lys Ser Ser Phe Phe Asn Ser Ala Ser Asp Val Phe Thr
                515                 520                 525

TAT GGG GTA GGA ACA GAT GTC CTC TAT AAC TTT ATC AAC GAT AAA GCC    1917
Tyr Gly Val Gly Thr Asp Val Leu Tyr Asn Phe Ile Asn Asp Lys Ala
            530                 535                 540

ACC AAA AAC AAT AAG ATT TCT TTT GGG GTG TTT GGG GGG ATT GCG TTA    1965
Thr Lys Asn Asn Lys Ile Ser Phe Gly Val Phe Gly Gly Ile Ala Leu
        545                 550                 555

GCT GGC ACT TCG TGG CTT AAT TCT CAA TAC GTG AAT TTA GCG ACA TTC    2013
Ala Gly Thr Ser Trp Leu Asn Ser Gln Tyr Val Asn Leu Ala Thr Phe
    560                 565                 570

AAT AAT TTT TAC AGC GCT AAA ATG AAT GTG GCG AAT TTC CAA TTC TTA    2061
Asn Asn Phe Tyr Ser Ala Lys Met Asn Val Ala Asn Phe Gln Phe Leu
575                 580                 585                 590

TTC AAC TTG GGC TTG AGA ATG AAT CTC GCT AAA AAC AAA AAG AAA GCG    2109
Phe Asn Leu Gly Leu Arg Met Asn Leu Ala Lys Asn Lys Lys Lys Ala
                595                 600                 605

AGC GAT CAT GTA GCT CAG CAT GGC GTG GAA CTA GGC GTG AAG ATC CCT    2157
Ser Asp His Val Ala Gln His Gly Val Glu Leu Gly Val Lys Ile Pro
            610                 615                 620

ACG ATC AAC ACG AAT TAC TAT TCT TTG CTA GGC ACT CAA CTC CAA TAC    2205
Thr Ile Asn Thr Asn Tyr Tyr Ser Leu Leu Gly Thr Gln Leu Gln Tyr
        625                 630                 635

CGC AGG CTT TAT AGC GTG TAT TTG AAT TAT GTG TTT GCT TAC TAATATCTG  2256
Arg Arg Leu Tyr Ser Val Tyr Leu Asn Tyr Val Phe Ala Tyr
    640                 645                 650

TCTTTTTGTG AAACTCCCTT TTTAAGGGAT TTTTTTTGAA GCCTTTCTTT TTTTAAACCC  2316
TCTTTTTTGG GGGTCAAGCG TAAAATTCAC CCCTATCCCT TTAAGAAAAT AAAATAAAAG  2376
AAAATGCGTT TTATAACAAA ATAAGATCTA AAACAATAAA ACAAAAACCC A            2427
```

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
  (A) LENGTH: 672 amino acids
  (B) TYPE: amino acid
  (C) STRANDEDNESS: single
  (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal
(ix) FEATURE:

(A) NAME/KEY: Signal Sequence
(B) LOCATION: 1...20
(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

```
Met Lys Lys Thr Leu Leu Leu Ser Leu Ser Leu Ser Leu Ser Ser Ser
-20                 -15                 -10                  -5
Leu Leu Asn Ala Glu Asp Asn Gly Phe Phe Ile Ser Ala Gly Tyr Gln
                 1               5                  10
Ile Gly Glu Ala Ala Gln Met Val Lys Asn Thr Gly Glu Leu Lys Lys
        15                  20                  25
Leu Ser Asp Thr Tyr Glu Asn Leu Ser Asn Leu Leu Thr Asn Phe Asn
    30                  35                  40
Asn Leu Asn Gln Ala Val Thr Asn Ala Ser Ser Pro Ser Glu Ile Asn
45                  50                  55                  60
Ala Thr Ile Asp Asn Leu Lys Ala Asn Thr Gln Gly Leu Ile Gly Glu
                65                  70                  75
Lys Thr Asn Ser Pro Ala Tyr Gln Ala Val Tyr Leu Ala Leu Asn Ala
            80                  85                  90
Ala Val Gly Leu Trp Asn Val Ile Ala Tyr Asn Val Gln Cys Gly Pro
        95                  100                 105
Gly Lys Ser Gly Asp Gln Ser Val Ile Phe Asp Gly Gln Pro Gly His
    110                 115                 120
Asp Ser Arg Ser Ile Asn Cys Asn Leu Thr Gly Tyr Asn Asn Gly Val
125                 130                 135                 140
Ser Gly Pro Leu Ser Ile Asp Asn Phe Lys Thr Leu Asn Gln Ala Tyr
                145                 150                 155
Gln Thr Ile Gln Gln Ala Leu Lys Gln Asp Ser Gly Phe Pro Val Leu
            160                 165                 170
Asp Ser Lys Gly Lys Gln Val Thr Ile Lys Ile Thr Thr Gln Thr Asn
        175                 180                 185
Gly Ala Asn Lys Ser Glu Thr Thr Thr Thr Thr Thr Thr Thr Asn Asp
    190                 195                 200
Ala Gln Thr Leu Leu Gln Glu Ala Ser Lys Met Ile Ser Val Leu Thr
205                 210                 215                 220
Thr Asn Cys Pro Trp Val Asn Thr Ala His Asn Ser Asn Gly Gly Ala
                225                 230                 235
Pro Trp Asn Leu Asn Thr Thr Gly Asn Val Cys Gln Val Phe Ala Thr
            240                 245                 250
Glu Phe Ser Ala Val Thr Ser Met Ile Lys Asn Ala Gln Glu Ile Val
        255                 260                 265
Thr Gln Ala Gln Ser Leu Asn Asn Pro Gln Ser Asn Gln Asn Ala Pro
    270                 275                 280
Lys Asp Phe Asn Pro Tyr Thr Ser Ala Asp Arg Ala Phe Ala Gln Asn
285                 290                 295                 300
Met Leu Asn His Ala Gln Ala Gln Ala Lys Met Leu Glu Leu Ala Asp
                305                 310                 315
Gln Met Lys Lys Asp Leu Asn Thr Ile Pro Lys Gln Phe Ile Thr Asn
            320                 325                 330
Tyr Leu Ala Ala Cys Arg Asn Gly Gly Gly Thr Leu Pro Asp Ala Gly
        335                 340                 345
Val Thr Ser Asn Thr Trp Gly Ala Gly Cys Ala Tyr Val Glu Glu Thr
    350                 355                 360
Ile Thr Ala Leu Asn Asn Ser Leu Ala His Phe Gly Thr Gln Ala Asp
```

```
365                 370                 375                 380
Gln Ile Lys Gln Ser Glu Leu Leu Ala Arg Thr Ile Leu Asp Phe Arg
                385                 390                 395
Gly Ser Leu Lys Asp Leu Asn Asn Thr Tyr Asn Ser Ile Thr Thr Thr
            400                 405                 410
Ala Ser Asn Thr Pro Asn Ser Pro Phe Leu Lys Asn Leu Ile Ser Gln
        415                 420                 425
Ser Thr Asn Pro Asn Asn Pro Gly Gly Leu Gln Ala Val Tyr Gln Val
    430                 435                 440
Asn Gln Ser Ala Tyr Ser Gln Leu Leu Ser Ala Thr Gln Glu Leu Gly
445                 450                 455                 460
His Asn Pro Phe Arg Arg Val Gly Leu Ile Ser Ser Gln Thr Asn Asn
                465                 470                 475
Gly Ala Met Asn Gly Ile Gly Val Gln Ile Gly Tyr Lys Gln Phe Phe
            480                 485                 490
Gly Glu Lys Arg Arg Trp Gly Leu Arg Tyr Tyr Gly Phe Phe Asp Tyr
        495                 500                 505
Asn His Ala Tyr Ile Lys Ser Ser Phe Phe Asn Ser Ala Ser Asp Val
    510                 515                 520
Phe Thr Tyr Gly Val Gly Thr Asp Val Leu Tyr Asn Phe Ile Asn Asp
525                 530                 535                 540
Lys Ala Thr Lys Asn Asn Lys Ile Ser Phe Gly Val Phe Gly Gly Ile
                545                 550                 555
Ala Leu Ala Gly Thr Ser Trp Leu Asn Ser Gln Tyr Val Asn Leu Ala
            560                 565                 570
Thr Phe Asn Asn Phe Tyr Ser Ala Lys Met Asn Val Ala Asn Phe Gln
        575                 580                 585
Phe Leu Phe Asn Leu Gly Leu Arg Met Asn Leu Ala Lys Asn Lys Lys
    590                 595                 600
Lys Ala Ser Asp His Val Ala Gln His Gly Val Glu Leu Gly Val Lys
605                 610                 615                 620
Ile Pro Thr Ile Asn Thr Asn Tyr Tyr Ser Leu Leu Gly Thr Gln Leu
                625                 630                 635
Gln Tyr Arg Arg Leu Tyr Ser Val Tyr Leu Asn Tyr Val Phe Ala Tyr
            640                 645                 650
```

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2429 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 205...2277
(D) OTHER INFORMATION:

(A) NAME/KEY: Signal Sequence
(B) LOCATION: 205...259
(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

```
TGAAAGAAGA CTGATTAGTC TTTCTTTTAG GGGCGATTCA AGCCTTAAAA GCCGGGTCAA      60
AATCCCCATT TTTCCCAATT TTTACAAAAA AAAAAAAAAC AAAATCTCTA AAATTTAGAG     120
CTAAAATTAG CCATAAAATT CCATTTATTG CTTATAATAT GAAGTTTCTT TGTATCAAAG     180
AAAAATCTAT TAAAAGGAGA AAAC ATG AAA AAA TCC CTC TTA CTC TCT CTT         231
                           Met Lys Lys Ser Leu Leu Leu Ser Leu
                                              -15                 -10

TCT CTC ATC GCT TCC TTA TCA AGA GCT GAA GAT GAC GGA TTT TAT ACG       279
Ser Leu Ile Ala Ser Leu Ser Arg Ala Glu Asp Asp Gly Phe Tyr Thr
                -5                   1               5

AGT GTG GGC TAT CAG ATC GGT GAA GCG GTC CAA CAA GTG AAA AAC ACA       327
Ser Val Gly Tyr Gln Ile Gly Glu Ala Val Gln Gln Val Lys Asn Thr
        10                  15                  20

GGA GCA TTG CAA AAT CTT GCA GAC AGA TAC GAT AAC TTA AAC AAC CTT       375
Gly Ala Leu Gln Asn Leu Ala Asp Arg Tyr Asp Asn Leu Asn Asn Leu
    25                  30                  35

TTA AAC CAA TAC AAT TAT TTA AAT TCC TTA GTC AAT TTA GCC AGC ACG       423
Leu Asn Gln Tyr Asn Tyr Leu Asn Ser Leu Val Asn Leu Ala Ser Thr
40                  45                  50                  55

CCG AGC GCG ATC ACC GGT GCG ATT GAT AAT TTA AGC TCA AGC GCG ATT       471
Pro Ser Ala Ile Thr Gly Ala Ile Asp Asn Leu Ser Ser Ser Ala Ile
                60                  65                  70

AAC CTC ACT AGC GCC ACC ACC ACT TCC CCC GCC TAT CAA GCT GTG GCT       519
Asn Leu Thr Ser Ala Thr Thr Thr Ser Pro Ala Tyr Gln Ala Val Ala
            75                  80                  85

TTA GCG CTC AAT GCC GCT GTG GGC ATG TGG CAA GTC ATA GCC CTT TTT       567
Leu Ala Leu Asn Ala Ala Val Gly Met Trp Gln Val Ile Ala Leu Phe
        90                  95                  100

ATT GGC TGT GGC CCT GGC CCT ACC AAT AAT CAA AGC TAT CAA TCG TTT       615
Ile Gly Cys Gly Pro Gly Pro Thr Asn Asn Gln Ser Tyr Gln Ser Phe
    105                 110                 115

GGT AAC ACA CCA GCC CTT AAT GGG ACC ACC ACC ACT TGC AAT CAA GCA       663
Gly Asn Thr Pro Ala Leu Asn Gly Thr Thr Thr Thr Cys Asn Gln Ala
120                 125                 130                 135

TAT GGG ACA GGC CCT AAT GGC ATC CTA TCT ATT GAT GAA TAC CAA AAA       711
Tyr Gly Thr Gly Pro Asn Gly Ile Leu Ser Ile Asp Glu Tyr Gln Lys
                140                 145                 150

CTC AAC CAA GCT TAT CAG ATC ATC CAA ACC GCT TTA AAC CAA AAT CAA       759
Leu Asn Gln Ala Tyr Gln Ile Ile Gln Thr Ala Leu Asn Gln Asn Gln
            155                 160                 165

GGG GGT GGG ATG CCT GCC TTG AAT GAC ACC ACC AAA ACA GGG GTA GTC       807
Gly Gly Gly Met Pro Ala Leu Asn Asp Thr Thr Lys Thr Gly Val Val
```

```
        170                 175                 180

AAC ATA CAA CAA ACC AAT TAT AGG ACC ACC ACA CAA AAC AAT ATC ATA     855
Asn Ile Gln Gln Thr Asn Tyr Arg Thr Thr Thr Gln Asn Asn Ile Ile
    185                 190                 195

GAG CAT TAT TAT ACA GAG AAT GGG AAA GAG ATC CCA GTC TCT TAT TCA     903
Glu His Tyr Tyr Thr Glu Asn Gly Lys Glu Ile Pro Val Ser Tyr Ser
200                 205                 210                 215

GGC GGA TCA TCA TTC TCG CCT ACA ATA CAA TTG ACA TAC CAT AAT AAC     951
Gly Gly Ser Ser Phe Ser Pro Thr Ile Gln Leu Thr Tyr His Asn Asn
                220                 225                 230

GCT GAA AAC CTT TTG CAA CAA GCC GCC ACT ATC ATG CAA GTC CTT ATT     999
Ala Glu Asn Leu Leu Gln Gln Ala Ala Thr Ile Met Gln Val Leu Ile
            235                 240                 245

ACT CAA AAG CCG CAT GTG CAA ACG AGC AAT GGC GGT AAA GCG TGG GGG    1047
Thr Gln Lys Pro His Val Gln Thr Ser Asn Gly Gly Lys Ala Trp Gly
        250                 255                 260

TTG AGT TCT ACG CCT GGG AAT GTG ATG GAT ATT TTT GGT CCT TCT TTT    1095
Leu Ser Ser Thr Pro Gly Asn Val Met Asp Ile Phe Gly Pro Ser Phe
    265                 270                 275

AAC GCT ATT AAT GAG ATG ATT AAA AAC GCT CAA ACA GCC CTA GCA AAA    1143
Asn Ala Ile Asn Glu Met Ile Lys Asn Ala Gln Thr Ala Leu Ala Lys
280                 285                 290                 295

ACC CAA CAG CTT AAC GCT AAT GAA AAC GCC CAA ATC ACG CAA CCC AAC    1191
Thr Gln Gln Leu Asn Ala Asn Glu Asn Ala Gln Ile Thr Gln Pro Asn
                300                 305                 310

AAT TTC AAC CCC TAC ACC TCT AAA GAC AAA GGG TTC GCT CAA GAA ATG    1239
Asn Phe Asn Pro Tyr Thr Ser Lys Asp Lys Gly Phe Ala Gln Glu Met
            315                 320                 325

CTC AAT AGA GCT GAA GCT CAA GCA GAG ATT TTA AAT TTA GCT AAG CAA    1287
Leu Asn Arg Ala Glu Ala Gln Ala Glu Ile Leu Asn Leu Ala Lys Gln
        330                 335                 340

GTA GCG AAC AAT TTC CAC AGC ATT CAA GGG CCT ATT CAA GGG GAT TTA    1335
Val Ala Asn Asn Phe His Ser Ile Gln Gly Pro Ile Gln Gly Asp Leu
    345                 350                 355

GAA GAA TGT AAA GCA GGA TCG GCT GGC GTG ATC ACT AAT AAC ACT TGG    1383
Glu Glu Cys Lys Ala Gly Ser Ala Gly Val Ile Thr Asn Asn Thr Trp
360                 365                 370                 375

GGT TCA GGT TGC GCG TTT GTG AAA GAA ACT TTA AAC TCT TTA GAG CAA    1431
Gly Ser Gly Cys Ala Phe Val Lys Glu Thr Leu Asn Ser Leu Glu Gln
                380                 385                 390

CAC ACC GCT TAT TAC GGC AAC CAG GTC AAT CAG GAT AGG GCT TTG GCT    1479
His Thr Ala Tyr Tyr Gly Asn Gln Val Asn Gln Asp Arg Ala Leu Ala
            395                 400                 405
```

```
CAA ACC ATT TTG AAT TTT AAA GAA GCC CTT AAC ACC CTG AAT AAA GAC    1527
Gln Thr Ile Leu Asn Phe Lys Glu Ala Leu Asn Thr Leu Asn Lys Asp
        410                 415                 420

TCA AAA GCG ATC AAT AGC GGT ATC TCC AAC TTG CCT AAC GCT AAA TCT    1575
Ser Lys Ala Ile Asn Ser Gly Ile Ser Asn Leu Pro Asn Ala Lys Ser
    425                 430                 435

CTT CAA AAC ATG ACG CAT GCC ACT CAA AAC CCT AAT TCC CCA GAA GGT    1623
Leu Gln Asn Met Thr His Ala Thr Gln Asn Pro Asn Ser Pro Glu Gly
440                 445                 450                 455

CTG CTC ACT TAT TCT TTG GAT TCA AGC AAA TAC AAC CAG CTC CAA ACC    1671
Leu Leu Thr Tyr Ser Leu Asp Ser Ser Lys Tyr Asn Gln Leu Gln Thr
                460                 465                 470

ATC GCG CAA GAA TTG GGC AAA AAC CCT TTC AGG CGC TTT GGC GTG ATT    1719
Ile Ala Gln Glu Leu Gly Lys Asn Pro Phe Arg Arg Phe Gly Val Ile
            475                 480                 485

GAC TTT CAA AAC AAC AAC GGC GCA ATG AAC GGG ATC GGC GTG CAA GTG    1767
Asp Phe Gln Asn Asn Asn Gly Ala Met Asn Gly Ile Gly Val Gln Val
        490                 495                 500

GGT TAT AAA CAA TTC TTT GGT AAA AAA AGG AAT TGG GGG TTA AGG TAT    1815
Gly Tyr Lys Gln Phe Phe Gly Lys Lys Arg Asn Trp Gly Leu Arg Tyr
    505                 510                 515

TAT GGT TTC TTT GAT TAT AAC CAT GCT TAT ATC AAA TCT AAT TTT TTC    1863
Tyr Gly Phe Phe Asp Tyr Asn His Ala Tyr Ile Lys Ser Asn Phe Phe
520                 525                 530                 535

AAC TCC GCT TCT GAT GTG TGG ACT TAT GGG GTG GGT ATG GAC GCT CTC    1911
Asn Ser Ala Ser Asp Val Trp Thr Tyr Gly Val Gly Met Asp Ala Leu
                540                 545                 550

TAT AAC TTC ATC AAC GAT AAA AAC ACC AAC TTT TTA GGC AAG AAC AAC    1959
Tyr Asn Phe Ile Asn Asp Lys Asn Thr Asn Phe Leu Gly Lys Asn Asn
            555                 560                 565

AAG CTT TCA GTA GGG CTT TTT GGA GGC TTT GCG TTA GCC GGG ACT TCG    2007
Lys Leu Ser Val Gly Leu Phe Gly Gly Phe Ala Leu Ala Gly Thr Ser
        570                 575                 580

TGG CTT AAT TCC CAA CAA GTG AAT TTG ACC ATG ATG AAT GGC ATT TAT    2055
Trp Leu Asn Ser Gln Gln Val Asn Leu Thr Met Met Asn Gly Ile Tyr
    585                 590                 595

AAC GCT AAT GTC AGC ACT TCT AAC TTC CAA TTT TTG TTT GAT TTA GGC    2103
Asn Ala Asn Val Ser Thr Ser Asn Phe Gln Phe Leu Phe Asp Leu Gly
600                 605                 610                 615

TTG AGA ATG AAC CTC GCT AGG CCT AAG AAA AAA GAC AGC GAT CAT GCC    2151
Leu Arg Met Asn Leu Ala Arg Pro Lys Lys Lys Asp Ser Asp His Ala
                620                 625                 630

GCT CAG CAT GGC ATT GAA CTA GGT TTT AAG ATC CCC ACG ATC AAC ACC    2199
```

```
Ala Gln His Gly Ile Glu Leu Gly Phe Lys Ile Pro Thr Ile Asn Thr
            635                 640                 645

AAC TAT TAT TCT TTC ATG GGC GCT AAA CTA GAA TAC AGA AGG ATG TAT    2247
Asn Tyr Tyr Ser Phe Met Gly Ala Lys Leu Glu Tyr Arg Arg Met Tyr
        650                 655                 660

AGC CTT TTT CTC AAT TAT GTG TTT GCT TAC TAAAAATTCT TTTTGAACCC CTC  2300
Ser Leu Phe Leu Asn Tyr Val Phe Ala Tyr
    665                 670

TTTTTTTGGG GGAGTGTTGC AAAAATGCCC CCCTATTTGC TTGTGAGTTT TGGTTAAAAT  2360
TTTAGTTACC CACGCTTAAA AAGCGCCAAG CCTTTTACAC ACAACTCCTT TAATTTTGTT  2420
TTTAAGAAA                                                          2429
```

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 691 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal
(ix) FEATURE:

(A) NAME/KEY: Signal Sequence
(B) LOCATION: 1...18
(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

```
Met Lys Lys Ser Leu Leu Leu Ser Leu Ser Leu Ile Ala Ser Leu Ser
           -15                 -10                 -5
Arg Ala Glu Asp Asp Gly Phe Tyr Thr Ser Val Gly Tyr Gln Ile Gly
         1                   5                  10
Glu Ala Val Gln Gln Val Lys Asn Thr Gly Ala Leu Gln Asn Leu Ala
15                  20                  25                  30
Asp Arg Tyr Asp Asn Leu Asn Asn Leu Leu Asn Gln Tyr Asn Tyr Leu
                35                  40                  45
Asn Ser Leu Val Asn Leu Ala Ser Thr Pro Ser Ala Ile Thr Gly Ala
            50                  55                  60
Ile Asp Asn Leu Ser Ser Ser Ala Ile Asn Leu Thr Ser Ala Thr Thr
        65                  70                  75
Thr Ser Pro Ala Tyr Gln Ala Val Ala Leu Ala Leu Asn Ala Ala Val
    80                  85                  90
Gly Met Trp Gln Val Ile Ala Leu Phe Ile Gly Cys Gly Pro Gly Pro
95                  100                 105                 110
Thr Asn Asn Gln Ser Tyr Gln Ser Phe Gly Asn Thr Pro Ala Leu Asn
                115                 120                 125
Gly Thr Thr Thr Thr Cys Asn Gln Ala Tyr Gly Thr Gly Pro Asn Gly
            130                 135                 140
Ile Leu Ser Ile Asp Glu Tyr Gln Lys Leu Asn Gln Ala Tyr Gln Ile
        145                 150                 155
```

```
Ile Gln Thr Ala Leu Asn Gln Asn Gln Gly Gly Gly Met Pro Ala Leu
    160                 165                 170
Asn Asp Thr Thr Lys Thr Gly Val Val Asn Ile Gln Gln Thr Asn Tyr
175                 180                 185                 190
Arg Thr Thr Thr Gln Asn Asn Ile Ile Glu His Tyr Tyr Thr Glu Asn
                195                 200                 205
Gly Lys Glu Ile Pro Val Ser Tyr Ser Gly Gly Ser Ser Phe Ser Pro
            210                 215                 220
Thr Ile Gln Leu Thr Tyr His Asn Asn Ala Glu Asn Leu Leu Gln Gln
        225                 230                 235
Ala Ala Thr Ile Met Gln Val Leu Ile Thr Gln Lys Pro His Val Gln
    240                 245                 250
Thr Ser Asn Gly Gly Lys Ala Trp Gly Leu Ser Ser Thr Pro Gly Asn
255                 260                 265                 270
Val Met Asp Ile Phe Gly Pro Ser Phe Asn Ala Ile Asn Glu Met Ile
                275                 280                 285
Lys Asn Ala Gln Thr Ala Leu Ala Lys Thr Gln Gln Leu Asn Ala Asn
            290                 295                 300
Glu Asn Ala Gln Ile Thr Gln Pro Asn Asn Phe Asn Pro Tyr Thr Ser
        305                 310                 315
Lys Asp Lys Gly Phe Ala Gln Glu Met Leu Asn Arg Ala Glu Ala Gln
    320                 325                 330
Ala Glu Ile Leu Asn Leu Ala Lys Gln Val Ala Asn Asn Phe His Ser
335                 340                 345                 350
Ile Gln Gly Pro Ile Gln Gly Asp Leu Glu Glu Cys Lys Ala Gly Ser
                355                 360                 365
Ala Gly Val Ile Thr Asn Asn Thr Trp Gly Ser Gly Cys Ala Phe Val
            370                 375                 380
Lys Glu Thr Leu Asn Ser Leu Glu Gln His Thr Ala Tyr Tyr Gly Asn
        385                 390                 395
Gln Val Asn Gln Asp Arg Ala Leu Ala Gln Thr Ile Leu Asn Phe Lys
    400                 405                 410
Glu Ala Leu Asn Thr Leu Asn Lys Asp Ser Lys Ala Ile Asn Ser Gly
415                 420                 425                 430
Ile Ser Asn Leu Pro Asn Ala Lys Ser Leu Gln Asn Met Thr His Ala
                435                 440                 445
Thr Gln Asn Pro Asn Ser Pro Glu Gly Leu Leu Thr Tyr Ser Leu Asp
            450                 455                 460
Ser Ser Lys Tyr Asn Gln Leu Gln Thr Ile Ala Gln Glu Leu Gly Lys
        465                 470                 475
Asn Pro Phe Arg Arg Phe Gly Val Ile Asp Phe Gln Asn Asn Asn Gly
    480                 485                 490
Ala Met Asn Gly Ile Gly Val Gln Val Gly Tyr Lys Gln Phe Phe Gly
495                 500                 505                 510
Lys Lys Arg Asn Trp Gly Leu Arg Tyr Tyr Gly Phe Phe Asp Tyr Asn
                515                 520                 525
His Ala Tyr Ile Lys Ser Asn Phe Phe Asn Ser Ala Ser Asp Val Trp
            530                 535                 540
Thr Tyr Gly Val Gly Met Asp Ala Leu Tyr Asn Phe Ile Asn Asp Lys
        545                 550                 555
Asn Thr Asn Phe Leu Gly Lys Asn Asn Lys Leu Ser Val Gly Leu Phe
    560                 565                 570
Gly Gly Phe Ala Leu Ala Gly Thr Ser Trp Leu Asn Ser Gln Gln Val
575                 580                 585                 590
Asn Leu Thr Met Met Asn Gly Ile Tyr Asn Ala Asn Val Ser Thr Ser
                595                 600                 605
Asn Phe Gln Phe Leu Phe Asp Leu Gly Leu Arg Met Asn Leu Ala Arg
```

```
            610                 615                 620
Pro Lys Lys Lys Asp Ser Asp His Ala Ala Gln His Gly Ile Glu Leu
        625                 630                 635
Gly Phe Lys Ile Pro Thr Ile Asn Thr Asn Tyr Tyr Ser Phe Met Gly
    640                 645                 650
Ala Lys Leu Glu Tyr Arg Arg Met Tyr Ser Leu Phe Leu Asn Tyr Val
655                 660                 665                 670
Phe Ala Tyr
```

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2270 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 130...2049
(D) OTHER INFORMATION:

(A) NAME/KEY: Signal Sequence
(B) LOCATION: 130...193
(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

```
ATTGAGCGCA TCAAAACACC CTAAAACTTT TTTGAAATCC AATAAATTTA TGTTATAATT     60
AAACGCATTG TAAATAAATT CTCATTTTGA TACATTTTTA CAATAAAACA TTACTTTAAG    120
GAACATCTT ATG AAA AAA ACG AAA AAA ACG ATT CTG CTT TCT CTA ACT CTC    171
          Met Lys Lys Thr Lys Lys Thr Ile Leu Leu Ser Leu Thr Leu
             -20                 -15                 -10

GCG GCG TCA TTG CTC CAT GCT GAA GAC AAC GGC GTT TTT TTA AGC GTG      219
Ala Ala Ser Leu Leu His Ala Glu Asp Asn Gly Val Phe Leu Ser Val
        -5                  1               5

GGT TAT CAA ATC GGT GAA GCG GTT CAA AAA GTG AAA AAC GCC GAC AAG      267
Gly Tyr Gln Ile Gly Glu Ala Val Gln Lys Val Lys Asn Ala Asp Lys
10                  15                  20                  25

GTG CAA AAA CTT TCA GAC ACT TAT GAA CAA TTA AGC CGG CTT TTA ACC      315
Val Gln Lys Leu Ser Asp Thr Tyr Glu Gln Leu Ser Arg Leu Leu Thr
                30                  35                  40

AAC GAT AAT GGC ACA AAC TCA AAG ACA AGC GCG CAA ATC AAC CAA GCG      363
Asn Asp Asn Gly Thr Asn Ser Lys Thr Ser Ala Gln Ile Asn Gln Ala
            45                  50                  55
```

```
GTT AAT AAT TTG AAC GAA CGC GCA AAA ACT TTA GCC GGT GGG ACA ACC      411
Val Asn Asn Leu Asn Glu Arg Ala Lys Thr Leu Ala Gly Gly Thr Thr
        60                  65                  70

AAT TCC CCT GCC TAT CAA GCC ACG CTT TTA GCG TTG AGA TCG GTG TTA      459
Asn Ser Pro Ala Tyr Gln Ala Thr Leu Leu Ala Leu Arg Ser Val Leu
    75                  80                  85

GGG CTA TGG AAT AGC ATG GGT TAT GCG GTC ATA TGC GGA GGT TAT ACC      507
Gly Leu Trp Asn Ser Met Gly Tyr Ala Val Ile Cys Gly Gly Tyr Thr
90                  95                  100                 105

AAA AGT CCA GGC GAA AAC AAT CAA AAA GAT TTC CAC TAC ACC GAT GAG      555
Lys Ser Pro Gly Glu Asn Asn Gln Lys Asp Phe His Tyr Thr Asp Glu
                110                 115                 120

AAT GGC AAT GGC ACT ACA ATC AAT TGC GGT GGG AGC ACA AAT AGT AAT      603
Asn Gly Asn Gly Thr Thr Ile Asn Cys Gly Gly Ser Thr Asn Ser Asn
            125                 130                 135

GGC ACT CAT AGT TCT AGT GGC ACA AAT ACA TTA AAA GCA GAC AAA AAT      651
Gly Thr His Ser Ser Ser Gly Thr Asn Thr Leu Lys Ala Asp Lys Asn
        140                 145                 150

GTT TCT CTA TCT ATT GAG CAA TAT GAA AAA ATC CAT GAA GCT TAT CAG      699
Val Ser Leu Ser Ile Glu Gln Tyr Glu Lys Ile His Glu Ala Tyr Gln
    155                 160                 165

ATT CTT TCA AAA GCT TTA AAA CAA GCC GGG CTT GCT CCT TTA AAT AGC      747
Ile Leu Ser Lys Ala Leu Lys Gln Ala Gly Leu Ala Pro Leu Asn Ser
170                 175                 180                 185

AAA GGG GAA AAG TTA GAA GCG CAT GTA ACC ACA TCA AAA CCA GAA AAT      795
Lys Gly Glu Lys Leu Glu Ala His Val Thr Thr Ser Lys Pro Glu Asn
                190                 195                 200

AAT AGT CAA ACT AAA ACG ACA ACT TCT GTT ATT GAT ACG ACT AAT GAT      843
Asn Ser Gln Thr Lys Thr Thr Thr Ser Val Ile Asp Thr Thr Asn Asp
            205                 210                 215

GCG CAA AAT CTT TTG ACT CAA GCG CAA ACG ATT GTC AAT ACC CTT AAA      891
Ala Gln Asn Leu Leu Thr Gln Ala Gln Thr Ile Val Asn Thr Leu Lys
        220                 225                 230

GAT TAT TGC CCC ATG TTG ATA GCG AAA TCT AGT AGT GAA AGT AGT GGC      939
Asp Tyr Cys Pro Met Leu Ile Ala Lys Ser Ser Ser Glu Ser Ser Gly
    235                 240                 245

GCA GCT ACT ACA AAC GCC CCT TCA TGG CAA ACA GCC GGT GGC GGC AAA      987
Ala Ala Thr Thr Asn Ala Pro Ser Trp Gln Thr Ala Gly Gly Gly Lys
250                 255                 260                 265

AAT TCA TGT GCG ACT TTT GGT GCG GAG TTT AGT GCC GCT TCA GAC ATG     1035
Asn Ser Cys Ala Thr Phe Gly Ala Glu Phe Ser Ala Ala Ser Asp Met
                270                 275                 280

ATT AAT AAT GCG CAA AAA ATC GTT CAA GAA ACC CAA CAA CTC AGC GCC     1083
```

```
Ile Asn Asn Ala Gln Lys Ile Val Gln Glu Thr Gln Gln Leu Ser Ala
            285                 290                 295

AAC CAA CCA AAA AAT ATC ACA CAA CCC CAT AAT CTC AAC CTT AAC ACC    1131
Asn Gln Pro Lys Asn Ile Thr Gln Pro His Asn Leu Asn Leu Asn Thr
        300                 305                 310

CCT AGC AGT CTT ACG GCT TTA GCT CAA AAA ATG CTC AAA AAT GCG CAA    1179
Pro Ser Ser Leu Thr Ala Leu Ala Gln Lys Met Leu Lys Asn Ala Gln
    315                 320                 325

TCT CAA GCA GAA ATT TTA AAA CTA GCC AAT CAA GTG GAG AGC GAT TTT    1227
Ser Gln Ala Glu Ile Leu Lys Leu Ala Asn Gln Val Glu Ser Asp Phe
330                 335                 340                 345

AAC AAA CTT TCT TCA GGC CAT CTT AAA GAC TAC ATA GGG AAA TGC GAT    1275
Asn Lys Leu Ser Ser Gly His Leu Lys Asp Tyr Ile Gly Lys Cys Asp
                350                 355                 360

GCG AGC GCT ATA AGC AGT GCG AAT ATG ACA ATG CAA AAT CAA AAG AAC    1323
Ala Ser Ala Ile Ser Ser Ala Asn Met Thr Met Gln Asn Gln Lys Asn
            365                 370                 375

AAT TGG GGG AAC GGG TGT GCT GGC GTG GAA GAA ACT CTG TCT TCA TTA    1371
Asn Trp Gly Asn Gly Cys Ala Gly Val Glu Glu Thr Leu Ser Ser Leu
        380                 385                 390

AAA ACA AGT GCC GCT GAT TTT AAC AAC CAA ACG CCA CAA ATC AAT CAA    1419
Lys Thr Ser Ala Ala Asp Phe Asn Asn Gln Thr Pro Gln Ile Asn Gln
    395                 400                 405

GCG CAA AAC CTA GCC AAC ACC CTT ATT CAA GAA CTT GGC AAC AAC CCT    1467
Ala Gln Asn Leu Ala Asn Thr Leu Ile Gln Glu Leu Gly Asn Asn Pro
410                 415                 420                 425

TTT AGG AAT ATG GGC ATG ATC GCT TCT TCA ACC ACG AAT AAC GGC GCC    1515
Phe Arg Asn Met Gly Met Ile Ala Ser Ser Thr Thr Asn Asn Gly Ala
                430                 435                 440

TTG AAT GGC CTT GGG GTG CAA GTG GGT TAT AAG CAA TTT TTT GGG GAA    1563
Leu Asn Gly Leu Gly Val Gln Val Gly Tyr Lys Gln Phe Phe Gly Glu
            445                 450                 455

AAG AAA AGA TGG GGG TTA AGG TAT TAT GGT TTC TTT GAT TAC AAC CAC    1611
Lys Lys Arg Trp Gly Leu Arg Tyr Tyr Gly Phe Phe Asp Tyr Asn His
        460                 465                 470

GCC TAT ATC AAA TCC AAT TTC TTT AAC TCG GCT TCT GAT GTG TGG ACT    1659
Ala Tyr Ile Lys Ser Asn Phe Phe Asn Ser Ala Ser Asp Val Trp Thr
    475                 480                 485

TAT GGG GTG GGC AGC GAT TTA TTG TTT AAT TTC ATC AAT GAT AAA AAC    1707
Tyr Gly Val Gly Ser Asp Leu Leu Phe Asn Phe Ile Asn Asp Lys Asn
490                 495                 500                 505

ACC AAC TTT TTA GGC AAG AAT AAC AAG ATT TCA GTG GGA TTT TTT GGA    1755
Thr Asn Phe Leu Gly Lys Asn Asn Lys Ile Ser Val Gly Phe Phe Gly
```

```
                510                 515                 520

GGT ATC GCC TTA GCA GGG ACT TCA TGG CTT AAT TCT CAA TTC GTG AAT    1803
Gly Ile Ala Leu Ala Gly Thr Ser Trp Leu Asn Ser Gln Phe Val Asn
            525                 530                 535

TTA AAA ACC ATC AGC AAT GTT TAT AGC GCT AAA GTG AAT ACG GCT AAC    1851
Leu Lys Thr Ile Ser Asn Val Tyr Ser Ala Lys Val Asn Thr Ala Asn
        540                 545                 550

TTC CAA TTT TTA TTC AAT TTG GGC TTG AGA ACC AAT CTC GCT AGA CCT    1899
Phe Gln Phe Leu Phe Asn Leu Gly Leu Arg Thr Asn Leu Ala Arg Pro
    555                 560                 565

AAG AAA AAA GAT AGT CAT CAT GCG GCT CAA CAT GGC ATG GAA TTG GGC    1947
Lys Lys Lys Asp Ser His His Ala Ala Gln His Gly Met Glu Leu Gly
570                 575                 580                 585

GTG AAA ATC CCT ACC ATT AAC ACG AAT TAT TAT TCT TTT CTA GAC ACT    1995
Val Lys Ile Pro Thr Ile Asn Thr Asn Tyr Tyr Ser Phe Leu Asp Thr
            590                 595                 600

AAA CTA GAA TAT CGA AGG CTT TAT AGC GTG TAT CTC AAT TAT GTG TTT    2043
Lys Leu Glu Tyr Arg Arg Leu Tyr Ser Val Tyr Leu Asn Tyr Val Phe
            605                 610                 615

GCC TAT TAAAAACCCT CTTTTTAAAA AAGGGGGGGC TTTAAAAAAC CTCTAAAGAT AA  2101
Ala Tyr


AAATTTTCAA AAAACAATCA TTAAACCCTA AAAAAGAAAT TTTAAGGTAT AATGCTTTCG  2161
CCATTTTTAA TTTTCCATGG CAAACTCCTT TTTAGAATTT ATCCCCATAA TCGCTCTTAT  2221
GGGGCGTTTG TTTTGCAACA ATCTTTTCGA AACTATCCAA CAAGCTTTA               2270
```

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 640 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal
(ix) FEATURE:

(A) NAME/KEY: Signal Sequence
(B) LOCATION: 1...21
(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

```
Met Lys Lys Thr Lys Lys Thr Ile Leu Leu Ser Leu Thr Leu Ala Ala
    -20                 -15                 -10
Ser Leu Leu His Ala Glu Asp Asn Gly Val Phe Leu Ser Val Gly Tyr
```

```
-5                  1               5                  10
Gln Ile Gly Glu Ala Val Gln Lys Val Lys Asn Ala Asp Lys Val Gln
            15                  20                  25
Lys Leu Ser Asp Thr Tyr Glu Gln Leu Ser Arg Leu Leu Thr Asn Asp
        30                  35                  40
Asn Gly Thr Asn Ser Lys Thr Ser Ala Gln Ile Asn Gln Ala Val Asn
    45                  50                  55
Asn Leu Asn Glu Arg Ala Lys Thr Leu Ala Gly Gly Thr Thr Asn Ser
60                  65                  70                  75
Pro Ala Tyr Gln Ala Thr Leu Leu Ala Leu Arg Ser Val Leu Gly Leu
                80                  85                  90
Trp Asn Ser Met Gly Tyr Ala Val Ile Cys Gly Gly Tyr Thr Lys Ser
            95                  100                 105
Pro Gly Glu Asn Asn Gln Lys Asp Phe His Tyr Thr Asp Glu Asn Gly
        110                 115                 120
Asn Gly Thr Thr Ile Asn Cys Gly Gly Ser Thr Asn Ser Asn Gly Thr
    125                 130                 135
His Ser Ser Ser Gly Thr Asn Thr Leu Lys Ala Asp Lys Asn Val Ser
140                 145                 150                 155
Leu Ser Ile Glu Gln Tyr Glu Lys Ile His Glu Ala Tyr Gln Ile Leu
                160                 165                 170
Ser Lys Ala Leu Lys Gln Ala Gly Leu Ala Pro Leu Asn Ser Lys Gly
            175                 180                 185
Glu Lys Leu Glu Ala His Val Thr Thr Ser Lys Pro Glu Asn Asn Ser
        190                 195                 200
Gln Thr Lys Thr Thr Thr Ser Val Ile Asp Thr Thr Asn Asp Ala Gln
    205                 210                 215
Asn Leu Leu Thr Gln Ala Gln Thr Ile Val Asn Thr Leu Lys Asp Tyr
220                 225                 230                 235
Cys Pro Met Leu Ile Ala Lys Ser Ser Ser Glu Ser Ser Gly Ala Ala
                240                 245                 250
Thr Thr Asn Ala Pro Ser Trp Gln Thr Ala Gly Gly Gly Lys Asn Ser
            255                 260                 265
Cys Ala Thr Phe Gly Ala Glu Phe Ser Ala Ala Ser Asp Met Ile Asn
        270                 275                 280
Asn Ala Gln Lys Ile Val Gln Glu Thr Gln Gln Leu Ser Ala Asn Gln
    285                 290                 295
Pro Lys Asn Ile Thr Gln Pro His Asn Leu Asn Leu Asn Thr Pro Ser
300                 305                 310                 315
Ser Leu Thr Ala Leu Ala Gln Lys Met Leu Lys Asn Ala Gln Ser Gln
                320                 325                 330
Ala Glu Ile Leu Lys Leu Ala Asn Gln Val Glu Ser Asp Phe Asn Lys
            335                 340                 345
Leu Ser Ser Gly His Leu Lys Asp Tyr Ile Gly Lys Cys Asp Ala Ser
        350                 355                 360
Ala Ile Ser Ser Ala Asn Met Thr Met Gln Asn Gln Lys Asn Asn Trp
    365                 370                 375
Gly Asn Gly Cys Ala Gly Val Glu Glu Thr Leu Ser Ser Leu Lys Thr
380                 385                 390                 395
Ser Ala Ala Asp Phe Asn Asn Gln Thr Pro Gln Ile Asn Gln Ala Gln
                400                 405                 410
Asn Leu Ala Asn Thr Leu Ile Gln Glu Leu Gly Asn Asn Pro Phe Arg
            415                 420                 425
Asn Met Gly Met Ile Ala Ser Ser Thr Thr Asn Asn Gly Ala Leu Asn
        430                 435                 440
Gly Leu Gly Val Gln Val Gly Tyr Lys Gln Phe Phe Gly Glu Lys Lys
    445                 450                 455
```

```
Arg Trp Gly Leu Arg Tyr Tyr Gly Phe Phe Asp Tyr Asn His Ala Tyr
460                 465                 470                 475
Ile Lys Ser Asn Phe Phe Asn Ser Ala Ser Asp Val Trp Thr Tyr Gly
                480                 485                 490
Val Gly Ser Asp Leu Leu Phe Asn Phe Ile Asn Asp Lys Asn Thr Asn
            495                 500                 505
Phe Leu Gly Lys Asn Asn Lys Ile Ser Val Gly Phe Phe Gly Gly Ile
        510                 515                 520
Ala Leu Ala Gly Thr Ser Trp Leu Asn Ser Gln Phe Val Asn Leu Lys
    525                 530                 535
Thr Ile Ser Asn Val Tyr Ser Ala Lys Val Asn Thr Ala Asn Phe Gln
540                 545                 550                 555
Phe Leu Phe Asn Leu Gly Leu Arg Thr Asn Leu Ala Arg Pro Lys Lys
                560                 565                 570
Lys Asp Ser His His Ala Ala Gln His Gly Met Glu Leu Gly Val Lys
            575                 580                 585
Ile Pro Thr Ile Asn Thr Asn Tyr Tyr Ser Phe Leu Asp Thr Lys Leu
        590                 595                 600
Glu Tyr Arg Arg Leu Tyr Ser Val Tyr Leu Asn Tyr Val Phe Ala Tyr
    605                 610                 615
```

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2248 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 173...2128
(D) OTHER INFORMATION:

(A) NAME/KEY: Signal Sequence
(B) LOCATION: 173...224
(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

```
TGGTTTTATC GTTACAAAAT TCAACATTTC AAAGATAAAT AAGTTAAAAT ACCCCAAAAT     60
CTTTTTTTTT TTTTTGAAAT CCAATCAATT TATAGTAAAA TTAGGTTCAT TGTAAATATA    120
TTATCACTTC ATGATATTCT TACAACAAAA ACATTACTTT AAGGAACATT TT ATG AAA    178
                                                          Met Lys

AAG ACA ATT CTG CTC TCT CTC TCT GCT TCA TCG CTC TTG CAC GCT GAA      226
Lys Thr Ile Leu Leu Ser Leu Ser Ala Ser Ser Leu Leu His Ala Glu
-15                 -10                 -5                  1

GAC AAC GGC TTT TTT GTG AGC GCC GGC TAT CAA ATC GGC GAA GCG GTG      274
```

```
Asp Asn Gly Phe Phe Val Ser Ala Gly Tyr Gln Ile Gly Glu Ala Val
                5                  10                  15

CAA ATG GTC AAA AAC ACC GGT GAA TTG AAA AAC TTG AAC GAA AAA TAC      322
Gln Met Val Lys Asn Thr Gly Glu Leu Lys Asn Leu Asn Glu Lys Tyr
        20                  25                  30

GAG CAA TTA AGC CAG TAT TTA AAT CAA GTG GCT TCG TTG AAG CAA AGC      370
Glu Gln Leu Ser Gln Tyr Leu Asn Gln Val Ala Ser Leu Lys Gln Ser
    35                  40                  45

ATT CAA AAC GCC AAC AAC ATT GAG CTG GTC AAT AGC TCT TTA AAC TAT      418
Ile Gln Asn Ala Asn Asn Ile Glu Leu Val Asn Ser Ser Leu Asn Tyr
50                  55                  60                  65

TTA AAA AGC TTT ACC AAC AAC AAC TAT AAC AGC ACC ACC CAA TCG CCC      466
Leu Lys Ser Phe Thr Asn Asn Asn Tyr Asn Ser Thr Thr Gln Ser Pro
                70                  75                  80

ATC TTT AAT GCC GTG CAA GCC GTT ATC ACT TCG GTA TTG GGT TTT TGG      514
Ile Phe Asn Ala Val Gln Ala Val Ile Thr Ser Val Leu Gly Phe Trp
            85                  90                  95

AGT CTT TAT GCG GGG AAT TAC TTC ACT TTT TTT GTG GGT AAA AAG GTG      562
Ser Leu Tyr Ala Gly Asn Tyr Phe Thr Phe Phe Val Gly Lys Lys Val
        100                 105                 110

GGT GAT AGT GGG CAA CCC GCT AGT GTC CAG GGT AAC CCT CCT TTT AAA      610
Gly Asp Ser Gly Gln Pro Ala Ser Val Gln Gly Asn Pro Pro Phe Lys
    115                 120                 125

ACG ATT ATA GAG AAC TGC TCA GGA ATT GAA AAC TGC GCT ATG GAT CAA      658
Thr Ile Ile Glu Asn Cys Ser Gly Ile Glu Asn Cys Ala Met Asp Gln
130                 135                 140                 145

ACC ACT TAT GAT AAG ATG AAA AAA CTC GCT GAA GAC CTC CAA GCG GCT      706
Thr Thr Tyr Asp Lys Met Lys Lys Leu Ala Glu Asp Leu Gln Ala Ala
                150                 155                 160

CAA ACA AAC TCT GCC ACT AAA GGC AAC AAT CTT TGC GCT TTA TCC GGG      754
Gln Thr Asn Ser Ala Thr Lys Gly Asn Asn Leu Cys Ala Leu Ser Gly
            165                 170                 175

TGT GCT GCA ACA GAC TCA ACA TCA AAC CCA CCA AAC TCA ACC GTG AGC      802
Cys Ala Ala Thr Asp Ser Thr Ser Asn Pro Pro Asn Ser Thr Val Ser
        180                 185                 190

AAC GCT CTT AAT TTG GCG CAA CAG CTT ATG GAT TTA ATC GCA AAC ACT      850
Asn Ala Leu Asn Leu Ala Gln Gln Leu Met Asp Leu Ile Ala Asn Thr
    195                 200                 205

AAA ACG GCT ATG ATG TGG AAA AAT ATC GTC ATC AGT GGC GTT TCA AAC      898
Lys Thr Ala Met Met Trp Lys Asn Ile Val Ile Ser Gly Val Ser Asn
210                 215                 220                 225

ACA TCC GGT GCT ATC ACA TCC ACT AAT TAC CCA ACG CAA TAC GCG GTG      946
Thr Ser Gly Ala Ile Thr Ser Thr Asn Tyr Pro Thr Gln Tyr Ala Val
```

```
                230                 235                 240

TTT AAC AAC ATT AAG GCG ATG ATA CCC ATT TTG CAA CAA GCG GTT ACG     994
Phe Asn Asn Ile Lys Ala Met Ile Pro Ile Leu Gln Gln Ala Val Thr
            245                 250                 255

CTT TCT CAA AGC AAC CAC ACC CTA TCT GCT AGC TTG CAA GCT CAA GCC    1042
Leu Ser Gln Ser Asn His Thr Leu Ser Ala Ser Leu Gln Ala Gln Ala
        260                 265                 270

ACA GGA TCT CAA ACA AAC CCT AAA TTC GCT AAA GAC ATC TAC ACT TTC    1090
Thr Gly Ser Gln Thr Asn Pro Lys Phe Ala Lys Asp Ile Tyr Thr Phe
    275                 280                 285

GCT CAA AAC CAA AAG CAA GTC ATC TCT TAC GCT CAA GAC ATT TTC AAC    1138
Ala Gln Asn Gln Lys Gln Val Ile Ser Tyr Ala Gln Asp Ile Phe Asn
290                 295                 300                 305

CTC TTT AAT TCT ATC CCT GCA GAG CAG TAT AAG TAT CTA GAG AAA GCT    1186
Leu Phe Asn Ser Ile Pro Ala Glu Gln Tyr Lys Tyr Leu Glu Lys Ala
                310                 315                 320

TAC TTG AAA ATA CCC AAT GCG GGT TCA ACG CCT ACT AAC CCT TAC AGA    1234
Tyr Leu Lys Ile Pro Asn Ala Gly Ser Thr Pro Thr Asn Pro Tyr Arg
            325                 330                 335

CAA GTG GTG AAT TTA AAC CAA GAA GTT CAG ACG ATT AAA AAC AAT GTG    1282
Gln Val Val Asn Leu Asn Gln Glu Val Gln Thr Ile Lys Asn Asn Val
        340                 345                 350

AGT TAT TAT GGT AAC CGG GTG GAT GCG GCT TTA AGC GTG GCT AGA GAT    1330
Ser Tyr Tyr Gly Asn Arg Val Asp Ala Ala Leu Ser Val Ala Arg Asp
    355                 360                 365

GTT TAT AAC CTA AAA TCC AAT CAA GCA GAA ATC GTA ACC GCC TAT AAC    1378
Val Tyr Asn Leu Lys Ser Asn Gln Ala Glu Ile Val Thr Ala Tyr Asn
370                 375                 380                 385

GAC GCT AAG ACT TTG AGC GAA GAG ATT TCT AAA CTC CCG CAC AAT CAA    1426
Asp Ala Lys Thr Leu Ser Glu Glu Ile Ser Lys Leu Pro His Asn Gln
                390                 395                 400

GTC AAT ACA AAA GAC ATT GTT ACA CTA CCT TAC GAT AAA AAC GCC CCA    1474
Val Asn Thr Lys Asp Ile Val Thr Leu Pro Tyr Asp Lys Asn Ala Pro
            405                 410                 415

GCA GCA GGC CAA TCC AAC TAC CAA ATC AAC CCA GAG CAG CAA TCC AAT    1522
Ala Ala Gly Gln Ser Asn Tyr Gln Ile Asn Pro Glu Gln Gln Ser Asn
        420                 425                 430

CTT AAC CAA GCT TTA GCA GCG ATG AGC AAT AAC CCC TTT AAA AAA GTG    1570
Leu Asn Gln Ala Leu Ala Ala Met Ser Asn Asn Pro Phe Lys Lys Val
    435                 440                 445

GGC ATG ATC AGC TCT CAA AAC AAT AAC GGC GCT TTG AAC GGG CTT GGC    1618
Gly Met Ile Ser Ser Gln Asn Asn Asn Gly Ala Leu Asn Gly Leu Gly
450                 455                 460                 465
```

```
GTG CAA GTG GGT TAT AAG CAA TTC TTT GGC GAA AGC AAA AGA TGG GGG    1666
Val Gln Val Gly Tyr Lys Gln Phe Phe Gly Glu Ser Lys Arg Trp Gly
                470                 475                 480

TTA AGG TAT TAC GGA TTC TTT GAT TAC AAC CAC GGC TAC ATC AAA TCC    1714
Leu Arg Tyr Tyr Gly Phe Phe Asp Tyr Asn His Gly Tyr Ile Lys Ser
            485                 490                 495

AGC TTC TTT AAC TCT TCT TCT GAT ATA TGG ACT TAT GGC GGT GGG AGC    1762
Ser Phe Phe Asn Ser Ser Ser Asp Ile Trp Thr Tyr Gly Gly Gly Ser
        500                 505                 510

GAT TTG TTA GTG AAT ATT ATC AAC GAT AGC ATC ACA AGA AAG AAC AAC    1810
Asp Leu Leu Val Asn Ile Ile Asn Asp Ser Ile Thr Arg Lys Asn Asn
    515                 520                 525

AAG CTC TCC GTG GGT CTT TTT GGA GGC ATC CAA CTA GCA GGG ACT ACA    1858
Lys Leu Ser Val Gly Leu Phe Gly Gly Ile Gln Leu Ala Gly Thr Thr
530                 535                 540                 545

TGG CTT AAT TCT CAA TAC GTG AAT TTA ACC GCG TTC AAT AAC CCT TAC    1906
Trp Leu Asn Ser Gln Tyr Val Asn Leu Thr Ala Phe Asn Asn Pro Tyr
                550                 555                 560

AGC GCG AAA GTC AAT GCT ACC AAT TTC CAA TTC TTG TTC AAT CTC GGC    1954
Ser Ala Lys Val Asn Ala Thr Asn Phe Gln Phe Leu Phe Asn Leu Gly
            565                 570                 575

TTG AGG ACG AAT CTC GCT ACA GCT AGG AAA AAA GAC AGC GAA CAT TCC    2002
Leu Arg Thr Asn Leu Ala Thr Ala Arg Lys Lys Asp Ser Glu His Ser
        580                 585                 590

GCG CAA CAT GGC ATT GAA TTG GGT ATT AAA ATC CCC ACC ATT ACC ACG    2050
Ala Gln His Gly Ile Glu Leu Gly Ile Lys Ile Pro Thr Ile Thr Thr
    595                 600                 605

AAT TAC TAT TCT TTT CTA GGC ACT CAA TTG CAA TAC AGA AGG CTC TAT    2098
Asn Tyr Tyr Ser Phe Leu Gly Thr Gln Leu Gln Tyr Arg Arg Leu Tyr
610                 615                 620                 625

AGC GTG TAT CTC AAT TAT GTG TTC GCT TAC TGAGTGATTC AAGCTCTCTT CTT  2151
Ser Val Tyr Leu Asn Tyr Val Phe Ala Tyr
                630                 635

TAAGGGGGTT TAGAAAAATC GCAACGCCAA GCTTTTTATC GTTGGTGATA AAATCTACAA   2211
AACTAACGGC GCGACAACAA ACCCTAACGC TACGCTC                            2248
```

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 652 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal

(ix) FEATURE:

(A) NAME/KEY: Signal Sequence
(B) LOCATION: 1...17
(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

```
Met Lys Lys Thr Ile Leu Leu Ser Leu Ser Ala Ser Ser Leu Leu His
        -15                 -10                 -5
Ala Glu Asp Asn Gly Phe Phe Val Ser Ala Gly Tyr Gln Ile Gly Glu
     1               5                  10                  15
Ala Val Gln Met Val Lys Asn Thr Gly Glu Leu Lys Asn Leu Asn Glu
                20                  25                  30
Lys Tyr Glu Gln Leu Ser Gln Tyr Leu Asn Gln Val Ala Ser Leu Lys
            35                  40                  45
Gln Ser Ile Gln Asn Ala Asn Asn Ile Glu Leu Val Asn Ser Ser Leu
        50                  55                  60
Asn Tyr Leu Lys Ser Phe Thr Asn Asn Asn Tyr Asn Ser Thr Thr Gln
    65                  70                  75
Ser Pro Ile Phe Asn Ala Val Gln Ala Val Ile Thr Ser Val Leu Gly
80                  85                  90                  95
Phe Trp Ser Leu Tyr Ala Gly Asn Tyr Phe Thr Phe Phe Val Gly Lys
                100                 105                 110
Lys Val Gly Asp Ser Gly Gln Pro Ala Ser Val Gln Gly Asn Pro Pro
            115                 120                 125
Phe Lys Thr Ile Ile Glu Asn Cys Ser Gly Ile Glu Asn Cys Ala Met
        130                 135                 140
Asp Gln Thr Thr Tyr Asp Lys Met Lys Lys Leu Ala Glu Asp Leu Gln
    145                 150                 155
Ala Ala Gln Thr Asn Ser Ala Thr Lys Gly Asn Asn Leu Cys Ala Leu
160                 165                 170                 175
Ser Gly Cys Ala Ala Thr Asp Ser Thr Ser Asn Pro Pro Asn Ser Thr
                180                 185                 190
Val Ser Asn Ala Leu Asn Leu Ala Gln Gln Leu Met Asp Leu Ile Ala
            195                 200                 205
Asn Thr Lys Thr Ala Met Met Trp Lys Asn Ile Val Ile Ser Gly Val
        210                 215                 220
Ser Asn Thr Ser Gly Ala Ile Thr Ser Thr Asn Tyr Pro Thr Gln Tyr
    225                 230                 235
Ala Val Phe Asn Asn Ile Lys Ala Met Ile Pro Ile Leu Gln Gln Ala
240                 245                 250                 255
Val Thr Leu Ser Gln Ser Asn His Thr Leu Ser Ala Ser Leu Gln Ala
                260                 265                 270
Gln Ala Thr Gly Ser Gln Thr Asn Pro Lys Phe Ala Lys Asp Ile Tyr
            275                 280                 285
Thr Phe Ala Gln Asn Gln Lys Gln Val Ile Ser Tyr Ala Gln Asp Ile
        290                 295                 300
Phe Asn Leu Phe Asn Ser Ile Pro Ala Glu Gln Tyr Lys Tyr Leu Glu
    305                 310                 315
Lys Ala Tyr Leu Lys Ile Pro Asn Ala Gly Ser Thr Pro Thr Asn Pro
320                 325                 330                 335
Tyr Arg Gln Val Val Asn Leu Asn Gln Glu Val Gln Thr Ile Lys Asn
                340                 345                 350
Asn Val Ser Tyr Tyr Gly Asn Arg Val Asp Ala Ala Leu Ser Val Ala
```

```
            355                 360                 365
Arg Asp Val Tyr Asn Leu Lys Ser Asn Gln Ala Glu Ile Val Thr Ala
        370                 375                 380
Tyr Asn Asp Ala Lys Thr Leu Ser Glu Glu Ile Ser Lys Leu Pro His
    385                 390                 395
Asn Gln Val Asn Thr Lys Asp Ile Val Thr Leu Pro Tyr Asp Lys Asn
400                 405                 410                 415
Ala Pro Ala Ala Gly Gln Ser Asn Tyr Gln Ile Asn Pro Glu Gln Gln
                420                 425                 430
Ser Asn Leu Asn Gln Ala Leu Ala Ala Met Ser Asn Asn Pro Phe Lys
            435                 440                 445
Lys Val Gly Met Ile Ser Ser Gln Asn Asn Asn Gly Ala Leu Asn Gly
        450                 455                 460
Leu Gly Val Gln Val Gly Tyr Lys Gln Phe Phe Gly Glu Ser Lys Arg
    465                 470                 475
Trp Gly Leu Arg Tyr Tyr Gly Phe Phe Asp Tyr Asn His Gly Tyr Ile
480                 485                 490                 495
Lys Ser Ser Phe Phe Asn Ser Ser Ser Asp Ile Trp Thr Tyr Gly Gly
                500                 505                 510
Gly Ser Asp Leu Leu Val Asn Ile Ile Asn Asp Ser Ile Thr Arg Lys
            515                 520                 525
Asn Asn Lys Leu Ser Val Gly Leu Phe Gly Gly Ile Gln Leu Ala Gly
        530                 535                 540
Thr Thr Trp Leu Asn Ser Gln Tyr Val Asn Leu Thr Ala Phe Asn Asn
    545                 550                 555
Pro Tyr Ser Ala Lys Val Asn Ala Thr Asn Phe Gln Phe Leu Phe Asn
560                 565                 570                 575
Leu Gly Leu Arg Thr Asn Leu Ala Thr Ala Arg Lys Lys Asp Ser Glu
                580                 585                 590
His Ser Ala Gln His Gly Ile Glu Leu Gly Ile Lys Ile Pro Thr Ile
            595                 600                 605
Thr Thr Asn Tyr Tyr Ser Phe Leu Gly Thr Gln Leu Gln Tyr Arg Arg
        610                 615                 620
Leu Tyr Ser Val Tyr Leu Asn Tyr Val Phe Ala Tyr
    625                 630                 635
```

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2161 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 122...2056
(D) OTHER INFORMATION:

(A) NAME/KEY: Signal Sequence
(B) LOCATION: 122...179
(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

```
CAAAAATCTT TTTTTTTTTT TTTTGAAATC CAATAAATTT ATGGTAAAGT TAAACATATT      60
GTAAATAAAT TTTAATTTCT ATTCATGTTT ACAATAAAAA AATTACTTTA AGGAACATTT     120
T ATG AAA AAG ACA ATT CTA CTC TCT CTC TCT CTC TCG CTT TCA TCG CTC     169
  Met Lys Lys Thr Ile Leu Leu Ser Leu Ser Leu Ser Leu Ser Ser Leu
                  -15                 -10                 -5

TTG CAC GCT GAA GAC AAC GGC TTT TTT GTG AGC GCC GGC TAT CAA ATC     217
Leu His Ala Glu Asp Asn Gly Phe Phe Val Ser Ala Gly Tyr Gln Ile
            1               5                  10

GGC GAA CGG GTG CAA ATG GTC AAA AAC ACC GGC GAA TTG AAA AAC TTG     265
Gly Glu Arg Val Gln Met Val Lys Asn Thr Gly Glu Leu Lys Asn Leu
    15                  20                  25

AAC GAA AAA TAC GAG CAA TTA AGC CAA TCT TTA GCC CAA CTG GCT TCG     313
Asn Glu Lys Tyr Glu Gln Leu Ser Gln Ser Leu Ala Gln Leu Ala Ser
30                  35                  40                  45

TTA AAA AAA AGC ATT CAA ACG GCG AAC AAC ATT CAG GCT GTC AAC AAT     361
Leu Lys Lys Ser Ile Gln Thr Ala Asn Asn Ile Gln Ala Val Asn Asn
                50                  55                  60

GCT TTA AGC GAT TTA AAA AGC TTT GCG AGT AAC AAC CAC ACA AAC AAA     409
Ala Leu Ser Asp Leu Lys Ser Phe Ala Ser Asn Asn His Thr Asn Lys
            65                  70                  75

GAA ACA TCG CCC ATC TAC AAC ACC GCG CAA GCT GTT ATC ACT TCA GTA     457
Glu Thr Ser Pro Ile Tyr Asn Thr Ala Gln Ala Val Ile Thr Ser Val
        80                  85                  90

TTG GCT TTT TGG AGT CTT TAT GCA GGG AAC GCT ACC AGT TTT CAT GTG     505
Leu Ala Phe Trp Ser Leu Tyr Ala Gly Asn Ala Thr Ser Phe His Val
    95                  100                 105

ACC GGT TTG AAT GAT GGA TCT AAT GCT CCT CTT GGA AGA ATC CAT CAA     553
Thr Gly Leu Asn Asp Gly Ser Asn Ala Pro Leu Gly Arg Ile His Gln
110                 115                 120                 125

GAT GGG AAC TGC ACA GGA TTA CAA CAA TGT TTT ATG AAT AAA GAA ACT     601
Asp Gly Asn Cys Thr Gly Leu Gln Gln Cys Phe Met Asn Lys Glu Thr
                130                 135                 140

TAT GAT AAA ATG AAA GCG CTT GCC GAA AAT CTC CAA AAA GCT CAA GGC     649
Tyr Asp Lys Met Lys Ala Leu Ala Glu Asn Leu Gln Lys Ala Gln Gly
            145                 150                 155

AAT CTC TGT GCC TTA TCA GAA TGC CCT AGC GAT CAA TTA AAT GGA AAC     697
Asn Leu Cys Ala Leu Ser Glu Cys Pro Ser Asp Gln Leu Asn Gly Asn
        160                 165                 170

AAT GGA AAC AAA ACT TCC ATG ACT AAA GCT CTT GAA ACC GCG CAA CAG     745
Asn Gly Asn Lys Thr Ser Met Thr Lys Ala Leu Glu Thr Ala Gln Gln
    175                 180                 185
```

```
CTT ATG GAT TTA ATC GCA AAC ACT AAA ACG GCT ATG ATG TGG AAA AAT      793
Leu Met Asp Leu Ile Ala Asn Thr Lys Thr Ala Met Met Trp Lys Asn
190                 195                 200                 205

ATC GTC ATC GCA GGT GTT ACA AAC AGA CCC GGT GGT GCT GGC GCT ATC      841
Ile Val Ile Ala Gly Val Thr Asn Arg Pro Gly Gly Ala Gly Ala Ile
                210                 215                 220

ACA TCC ACT GGT CCT GTA ACC GAC TAT GCG GTG TTT AAC AAC ATT AAG      889
Thr Ser Thr Gly Pro Val Thr Asp Tyr Ala Val Phe Asn Asn Ile Lys
            225                 230                 235

GCG ATG ATA CCC ATT TTG CAA CAA GCG GTT ACG CTT TCT CAA AGC AAC      937
Ala Met Ile Pro Ile Leu Gln Gln Ala Val Thr Leu Ser Gln Ser Asn
        240                 245                 250

CAC ACC CTA TCT GCT AGC TTG CAA GCT CAA GCC ACA GGA TCT CAA ACA      985
His Thr Leu Ser Ala Ser Leu Gln Ala Gln Ala Thr Gly Ser Gln Thr
    255                 260                 265

AAC CCT AAA TTC GCT AAA GAC ATC TAC ACT TTC GCT CAA AAC CAA AAG     1033
Asn Pro Lys Phe Ala Lys Asp Ile Tyr Thr Phe Ala Gln Asn Gln Lys
270                 275                 280                 285

CAA GTC ATC TCT TAC GCT CAA GAC ATT TTC AAC CTC TTT AAT TCT ATC     1081
Gln Val Ile Ser Tyr Ala Gln Asp Ile Phe Asn Leu Phe Asn Ser Ile
                290                 295                 300

CCT GCA GAG CAG TAT AAG TAT CTA GAG AAA GCT TAC TTG AAA ATA CCC     1129
Pro Ala Glu Gln Tyr Lys Tyr Leu Glu Lys Ala Tyr Leu Lys Ile Pro
            305                 310                 315

AAT GCG GGT TCA ACG CCT ACT AAC CCT TAC AGA CAA GTG GTG AAT TTA     1177
Asn Ala Gly Ser Thr Pro Thr Asn Pro Tyr Arg Gln Val Val Asn Leu
        320                 325                 330

AAC CAA GAA GTT CAG ACG ATT AAA AAC AAT GTG AGT TAT TAT GGT AAC     1225
Asn Gln Glu Val Gln Thr Ile Lys Asn Asn Val Ser Tyr Tyr Gly Asn
    335                 340                 345

CGG GTG GAT GCG GCT TTA AGC GTG GCT AGA GAT GTT TAT AAC CTA AAA     1273
Arg Val Asp Ala Ala Leu Ser Val Ala Arg Asp Val Tyr Asn Leu Lys
350                 355                 360                 365

TCC AAT CAA GCA GAA ATC GTA ACC GCC TAT AAC GAC GCT AAG ACT TTG     1321
Ser Asn Gln Ala Glu Ile Val Thr Ala Tyr Asn Asp Ala Lys Thr Leu
                370                 375                 380

AGC GAA GAG ATT TCT AAA CTC CCG CAC AAT CAA GTC AAT ACA AAA GAC     1369
Ser Glu Glu Ile Ser Lys Leu Pro His Asn Gln Val Asn Thr Lys Asp
            385                 390                 395

ATT GTT ACA CTA CCT TAC GAT AAA AAC GCC CCA GCA GCA GGC CAA TCC     1417
Ile Val Thr Leu Pro Tyr Asp Lys Asn Ala Pro Ala Ala Gly Gln Ser
        400                 405                 410

AAC TAC CAA ATC AAC CCA GAG CAG CAA TCC AAT CTT AAC CAA GCT TTA     1465
```

```
Asn Tyr Gln Ile Asn Pro Glu Gln Gln Ser Asn Leu Asn Gln Ala Leu
    415                 420                 425

GCA GCG ATG AGC AAT AAC CCC TTT AAA AAA GTG GGC ATG ATC AGC TCT    1513
Ala Ala Met Ser Asn Asn Pro Phe Lys Lys Val Gly Met Ile Ser Ser
430                 435                 440                 445

CAA AAC AAT AAC GGC GCT TTG AAC GGG CTT GGC GTG CAA GTG GGT TAT    1561
Gln Asn Asn Asn Gly Ala Leu Asn Gly Leu Gly Val Gln Val Gly Tyr
                450                 455                 460

AAG CAA TTC TTT GGC GAA AGC AAA AGA TGG GGG TTA AGG TAT TAC GGA    1609
Lys Gln Phe Phe Gly Glu Ser Lys Arg Trp Gly Leu Arg Tyr Tyr Gly
            465                 470                 475

TTC TTT GAT TAC AAC CAC GGC TAC ATC AAA TCC AGC TTC TTT AAC TCT    1657
Phe Phe Asp Tyr Asn His Gly Tyr Ile Lys Ser Ser Phe Phe Asn Ser
        480                 485                 490

TCT TCT GAT ATA TGG ACT TAT GGC GGT GGG AGC GAT TTG TTA GTG AAT    1705
Ser Ser Asp Ile Trp Thr Tyr Gly Gly Gly Ser Asp Leu Leu Val Asn
    495                 500                 505

ATT ATC AAC GAT AGC ATC ACA AGA AAG AAC AAC AAG CTC TCC GTG GGT    1753
Ile Ile Asn Asp Ser Ile Thr Arg Lys Asn Asn Lys Leu Ser Val Gly
510                 515                 520                 525

CTT TTT GGA GGC ATC CAA CTA GCA GGG ACT ACA TGG CTT AAT TCT CAA    1801
Leu Phe Gly Gly Ile Gln Leu Ala Gly Thr Thr Trp Leu Asn Ser Gln
                530                 535                 540

TAC GTG AAT TTA ACC GCG TTC AAT AAC CCT TAC AGC GCG AAA GTC AAT    1849
Tyr Val Asn Leu Thr Ala Phe Asn Asn Pro Tyr Ser Ala Lys Val Asn
            545                 550                 555

GCT ACC AAT TTC CAA TTC TTG TTC AAT CTC GGC TTG AGG ACG AAT CTC    1897
Ala Thr Asn Phe Gln Phe Leu Phe Asn Leu Gly Leu Arg Thr Asn Leu
        560                 565                 570

GCT ACA GCT AGG AAA AAA GAC AGC GAA CAT TCC GCG CAA CAT GGC ATT    1945
Ala Thr Ala Arg Lys Lys Asp Ser Glu His Ser Ala Gln His Gly Ile
    575                 580                 585

GAA TTG GGT ATT AAA ATC CCC ACC ATT ACC ACG AAT TAC TAT TCT TTT    1993
Glu Leu Gly Ile Lys Ile Pro Thr Ile Thr Thr Asn Tyr Tyr Ser Phe
590                 595                 600                 605

CTA GGC ACT CAA TTG CAA TAC AGA AGG CTC TAT AGC GTG TAT CTC AAT    2041
Leu Gly Thr Gln Leu Gln Tyr Arg Arg Leu Tyr Ser Val Tyr Leu Asn
                610                 615                 620

TAT GTG TTC GCT TAT TAAAAAATCT TCTTTTTAAA ATAGGGGGAG CTTCATCAAA T  2097
Tyr Val Phe Ala Tyr
            625

CTATTTTGAT AGTTATCAAT ATTTGATGAA AATAAAGTCA AAAACAAAAT AAACCAAATC   2157
ACCC                                                               2161
```

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 645 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal
(ix) FEATURE:

(A) NAME/KEY: Signal Sequence
(B) LOCATION: 1...19
(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

```
Met Lys Lys Thr Ile Leu Leu Ser Leu Ser Leu Ser Leu Ser Ser Leu
                -15                 -10                  -5
Leu His Ala Glu Asp Asn Gly Phe Phe Val Ser Ala Gly Tyr Gln Ile
            1               5                   10
Gly Glu Arg Val Gln Met Val Lys Asn Thr Gly Glu Leu Lys Asn Leu
    15                  20                  25
Asn Glu Lys Tyr Glu Gln Leu Ser Gln Ser Leu Ala Gln Leu Ala Ser
30                  35                  40                  45
Leu Lys Lys Ser Ile Gln Thr Ala Asn Asn Ile Gln Ala Val Asn Asn
                50                  55                  60
Ala Leu Ser Asp Leu Lys Ser Phe Ala Ser Asn Asn His Thr Asn Lys
            65                  70                  75
Glu Thr Ser Pro Ile Tyr Asn Thr Ala Gln Ala Val Ile Thr Ser Val
        80                  85                  90
Leu Ala Phe Trp Ser Leu Tyr Ala Gly Asn Ala Thr Ser Phe His Val
    95                  100                 105
Thr Gly Leu Asn Asp Gly Ser Asn Ala Pro Leu Gly Arg Ile His Gln
110                 115                 120                 125
Asp Gly Asn Cys Thr Gly Leu Gln Gln Cys Phe Met Asn Lys Glu Thr
                130                 135                 140
Tyr Asp Lys Met Lys Ala Leu Ala Glu Asn Leu Gln Lys Ala Gln Gly
            145                 150                 155
Asn Leu Cys Ala Leu Ser Glu Cys Pro Ser Asp Gln Leu Asn Gly Asn
        160                 165                 170
Asn Gly Asn Lys Thr Ser Met Thr Lys Ala Leu Glu Thr Ala Gln Gln
    175                 180                 185
Leu Met Asp Leu Ile Ala Asn Thr Lys Thr Ala Met Met Trp Lys Asn
190                 195                 200                 205
Ile Val Ile Ala Gly Val Thr Asn Arg Pro Gly Gly Ala Gly Ala Ile
                210                 215                 220
Thr Ser Thr Gly Pro Val Thr Asp Tyr Ala Val Phe Asn Asn Ile Lys
            225                 230                 235
Ala Met Ile Pro Ile Leu Gln Gln Ala Val Thr Leu Ser Gln Ser Asn
        240                 245                 250
His Thr Leu Ser Ala Ser Leu Gln Ala Gln Ala Thr Gly Ser Gln Thr
    255                 260                 265
Asn Pro Lys Phe Ala Lys Asp Ile Tyr Thr Phe Ala Gln Asn Gln Lys
```

```
270                 275                 280                 285
Gln Val Ile Ser Tyr Ala Gln Asp Ile Phe Asn Leu Phe Asn Ser Ile
                290                 295                 300
Pro Ala Glu Gln Tyr Lys Tyr Leu Glu Lys Ala Tyr Leu Lys Ile Pro
            305                 310                 315
Asn Ala Gly Ser Thr Pro Thr Asn Pro Tyr Arg Gln Val Val Asn Leu
        320                 325                 330
Asn Gln Glu Val Gln Thr Ile Lys Asn Asn Val Ser Tyr Tyr Gly Asn
    335                 340                 345
Arg Val Asp Ala Ala Leu Ser Val Ala Arg Asp Val Tyr Asn Leu Lys
350                 355                 360                 365
Ser Asn Gln Ala Glu Ile Val Thr Ala Tyr Asn Asp Ala Lys Thr Leu
                370                 375                 380
Ser Glu Glu Ile Ser Lys Leu Pro His Asn Gln Val Asn Thr Lys Asp
            385                 390                 395
Ile Val Thr Leu Pro Tyr Asp Lys Asn Ala Pro Ala Ala Gly Gln Ser
        400                 405                 410
Asn Tyr Gln Ile Asn Pro Glu Gln Gln Ser Asn Leu Asn Gln Ala Leu
    415                 420                 425
Ala Ala Met Ser Asn Asn Pro Phe Lys Lys Val Gly Met Ile Ser Ser
430                 435                 440                 445
Gln Asn Asn Asn Gly Ala Leu Asn Gly Leu Gly Val Gln Val Gly Tyr
                450                 455                 460
Lys Gln Phe Phe Gly Glu Ser Lys Arg Trp Gly Leu Arg Tyr Tyr Gly
            465                 470                 475
Phe Phe Asp Tyr Asn His Gly Tyr Ile Lys Ser Ser Phe Phe Asn Ser
        480                 485                 490
Ser Ser Asp Ile Trp Thr Tyr Gly Gly Gly Ser Asp Leu Leu Val Asn
    495                 500                 505
Ile Ile Asn Asp Ser Ile Thr Arg Lys Asn Asn Lys Leu Ser Val Gly
510                 515                 520                 525
Leu Phe Gly Gly Ile Gln Leu Ala Gly Thr Thr Trp Leu Asn Ser Gln
                530                 535                 540
Tyr Val Asn Leu Thr Ala Phe Asn Asn Pro Tyr Ser Ala Lys Val Asn
            545                 550                 555
Ala Thr Asn Phe Gln Phe Leu Phe Asn Leu Gly Leu Arg Thr Asn Leu
        560                 565                 570
Ala Thr Ala Arg Lys Lys Asp Ser Glu His Ser Ala Gln His Gly Ile
    575                 580                 585
Glu Leu Gly Ile Lys Ile Pro Thr Ile Thr Thr Asn Tyr Tyr Ser Phe
590                 595                 600                 605
Leu Gly Thr Gln Leu Gln Tyr Arg Arg Leu Tyr Ser Val Tyr Leu Asn
                610                 615                 620
Tyr Val Phe Ala Tyr
            625
```

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1799 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 185...1633
(D) OTHER INFORMATION:

(A) NAME/KEY: Signal Sequence
(B) LOCATION: 185...233
(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

```
TACTCAAAAC ATTTTTCACT ATCAAAAACC TTTTTTTTAA ATCCAAAAAA AAAGCAAAAT     60
TTCTTAATTT TTGCTCAATT TTATTAAAAA TTCAATAAAT TTATGGCACA ATTTAAACTT    120
ATTGTAAATA AAGTTTCAAT TTGATACGAT TTTACAAACA AAACATTACT TTAAGGAACA    180
TTTT ATG AAA AAA ACG ATT TTA CTT TCT CTT ATG GTT TCA TCG CTC CTC     229
     Met Lys Lys Thr Ile Leu Leu Ser Leu Met Val Ser Ser Leu Leu
         -15                 -10                 -5

GCT GAA AAT GAC GGC GTT TTT ATG AGC GTG GGC TAT CAA ATC GGC GAA      277
Ala Glu Asn Asp Gly Val Phe Met Ser Val Gly Tyr Gln Ile Gly Glu
 1               5                  10                  15

GCG GTT CAA CAA GTG AAA AAC ACC GGC GAA ATC CAA AAA GTC TCC AAC      325
Ala Val Gln Gln Val Lys Asn Thr Gly Glu Ile Gln Lys Val Ser Asn
                20                  25                  30

GCT TAC GAA AAT TTG AAC AAT CTT TTA ACC CGC TAT AAC GAA CTC AAA      373
Ala Tyr Glu Asn Leu Asn Asn Leu Leu Thr Arg Tyr Asn Glu Leu Lys
            35                  40                  45

CAA ACG GCC TCT AAC ACC AAT TCA AGT ACC GCT CAA GCG ATT GAT AAT      421
Gln Thr Ala Ser Asn Thr Asn Ser Ser Thr Ala Gln Ala Ile Asp Asn
        50                  55                  60

CTA AAA GAG AGC GCT AGC CGA TTG AAA ACG ACC CCC AAT AGC GCT AAT      469
Leu Lys Glu Ser Ala Ser Arg Leu Lys Thr Thr Pro Asn Ser Ala Asn
    65                  70                  75

CAA GCC GTG TCT TCA GCG CTC AGC TCT GCG GTA GCC ATG TGG CAA GTA      517
Gln Ala Val Ser Ser Ala Leu Ser Ser Ala Val Ala Met Trp Gln Val
80                  85                  90                  95

ATA GTC TCT AAT TTA GCC AAT AAC TCG CTA CCC ACT AGT GAA TAC AAC      565
Ile Val Ser Asn Leu Ala Asn Asn Ser Leu Pro Thr Ser Glu Tyr Asn
                100                 105                 110

AAA ATC AAT GCG ATT TCT CAA TCG CTC CAA AAC ACC CTA GAA AAT AAA      613
Lys Ile Asn Ala Ile Ser Gln Ser Leu Gln Asn Thr Leu Glu Asn Lys
            115                 120                 125

AAC AAT GAT CTT AAA ATT GAA AAT GAC TAC GAC CAT CTT TTA ACT CAA      661
Asn Asn Asp Leu Lys Ile Glu Asn Asp Tyr Asp His Leu Leu Thr Gln
        130                 135                 140
```

```
GCT AGC ACC ATT ATT AAT ACC CTT CAA AGC CAA TGC CCA GGC ATA GAC      709
Ala Ser Thr Ile Ile Asn Thr Leu Gln Ser Gln Cys Pro Gly Ile Asp
    145                 150                 155

GGA GGC AAT GGC AAA CCA TGG GGC ATT AAT GCA AGC GGG AAC GCA TGC      757
Gly Gly Asn Gly Lys Pro Trp Gly Ile Asn Ala Ser Gly Asn Ala Cys
160                 165                 170                 175

AAT ATT TTT GGC AAC ACC TTT AAC GCC ATC ACT AGC ATG ATA GAT AGC      805
Asn Ile Phe Gly Asn Thr Phe Asn Ala Ile Thr Ser Met Ile Asp Ser
                180                 185                 190

GCT AAA AAA GCC GCC GCA GAT GCC CGA AGA ACT GCC CCA GAA AGT CCA      853
Ala Lys Lys Ala Ala Ala Asp Ala Arg Arg Thr Ala Pro Glu Ser Pro
            195                 200                 205

AAC CAA CCA AGT GCG TTT AAC AAC GCT GAT TTC AAT AAA AAC CTT AAT      901
Asn Gln Pro Ser Ala Phe Asn Asn Ala Asp Phe Asn Lys Asn Leu Asn
        210                 215                 220

CAA GTC TCA AGC GTT ATT AAT GAC ACG ATC TCT TAC CTC AAA GGG GAC      949
Gln Val Ser Ser Val Ile Asn Asp Thr Ile Ser Tyr Leu Lys Gly Asp
    225                 230                 235

AAT TTA GCA ACC ATC TAC AAC ACC CTT CAA AAA ACG CCC GAT TCT AAA      997
Asn Leu Ala Thr Ile Tyr Asn Thr Leu Gln Lys Thr Pro Asp Ser Lys
240                 245                 250                 255

GGG TTT CAA AGT TTG GTG AGC CGA TCT AGC TAT AGT TAT TCC CTC AAC     1045
Gly Phe Gln Ser Leu Val Ser Arg Ser Ser Tyr Ser Tyr Ser Leu Asn
                260                 265                 270

GAA ACC CAA TAT TCT GAA TTC CAA ACT ACC ACC AAA GAG TTT GGC CAT     1093
Glu Thr Gln Tyr Ser Glu Phe Gln Thr Thr Thr Lys Glu Phe Gly His
            275                 280                 285

AAC CCT TTT AGA AGC GTG GGT TTA ATC AAC TCT CAA AGC AAT AAC GGA     1141
Asn Pro Phe Arg Ser Val Gly Leu Ile Asn Ser Gln Ser Asn Asn Gly
        290                 295                 300

GCG ATG AAT GGC GTG GGC GTG CAA TTA GGC TAT AAG CAA TTC TTT GGG     1189
Ala Met Asn Gly Val Gly Val Gln Leu Gly Tyr Lys Gln Phe Phe Gly
    305                 310                 315

AAA AAT AAA TTT TTT GGG ATC CGT TAT TAT GCC TTT TTT GAT TAC AAC     1237
Lys Asn Lys Phe Phe Gly Ile Arg Tyr Tyr Ala Phe Phe Asp Tyr Asn
320                 325                 330                 335

CAT GCC TAT ATC AAA TCC AAC TTT TTC AAC TCC GCT TCC AAT GTT TTC     1285
His Ala Tyr Ile Lys Ser Asn Phe Phe Asn Ser Ala Ser Asn Val Phe
                340                 345                 350

ACT TAT GGC GCA GGC AGT GAT CTT TTA TTG AAT TTC ATC AAT GGC GGA     1333
Thr Tyr Gly Ala Gly Ser Asp Leu Leu Leu Asn Phe Ile Asn Gly Gly
            355                 360                 365

TCC GAT AAA AAC CGC AAA GTC TCT TTT GGC ATT TTT GGA GGC ATC GCT     1381
```

```
Ser Asp Lys Asn Arg Lys Val Ser Phe Gly Ile Phe Gly Gly Ile Ala
        370                 375                 380

CTA GCA GGC ACG ACA TGG CTT AAT TCC CAA TTT ATG AAT TTA AAA ACC    1429
Leu Ala Gly Thr Thr Trp Leu Asn Ser Gln Phe Met Asn Leu Lys Thr
    385                 390                 395

ACC AAT AGC GCC TAC AGC GCT AAG ATC AAC AAC ACC AAT TTC CAA TTC    1477
Thr Asn Ser Ala Tyr Ser Ala Lys Ile Asn Asn Thr Asn Phe Gln Phe
400                 405                 410                 415

TTA TTC AAT ACT GGT TTA AGG CTT CAA GGG ATT CAC CAT GGC GTT GAA    1525
Leu Phe Asn Thr Gly Leu Arg Leu Gln Gly Ile His His Gly Val Glu
                420                 425                 430

TTA GGC GTG AAA ATC CCC ACC ATC AAC ACG AAT TAC TAT TCT TTC ATG    1573
Leu Gly Val Lys Ile Pro Thr Ile Asn Thr Asn Tyr Tyr Ser Phe Met
            435                 440                 445

GGC GCT AAA TTA GCA TAC CGA AGA CTT TAT AGC GTG TAT TTC AAT TAT    1621
Gly Ala Lys Leu Ala Tyr Arg Arg Leu Tyr Ser Val Tyr Phe Asn Tyr
        450                 455                 460

GTT TTG GCC TAT TGATATTGAA TCGGTTCTCA TTACTAATGA GGACAAAGCC AAACT  1678
Val Leu Ala Tyr
    465

TTTTGGCTCT CAATGAATAA CGGCATCATT TTACTTGACT TTTTACAAAA AACACACTAA   1738
AATTTCTTTT TCTTTTTTGA GCGAAATTCC AGATTAGCTC AGCGGTAGAG TAGGCGGCTG   1798
T                                                                   1799
```

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 483 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal
(ix) FEATURE:

(A) NAME/KEY: Signal Sequence
(B) LOCATION: 1...16
(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

```
Met Lys Lys Thr Ile Leu Leu Ser Leu Met Val Ser Ser Leu Leu Ala
    -15                 -10                 -5
Glu Asn Asp Gly Val Phe Met Ser Val Gly Tyr Gln Ile Gly Glu Ala
 1                5                  10                  15
Val Gln Gln Val Lys Asn Thr Gly Glu Ile Gln Lys Val Ser Asn Ala
            20                  25                  30
```

```
Tyr Glu Asn Leu Asn Asn Leu Leu Thr Arg Tyr Asn Glu Leu Lys Gln
        35                  40                  45
Thr Ala Ser Asn Thr Asn Ser Ser Thr Ala Gln Ala Ile Asp Asn Leu
    50                  55                  60
Lys Glu Ser Ala Ser Arg Leu Lys Thr Thr Pro Asn Ser Ala Asn Gln
65                  70                  75                  80
Ala Val Ser Ser Ala Leu Ser Ser Ala Val Ala Met Trp Gln Val Ile
                85                  90                  95
Val Ser Asn Leu Ala Asn Asn Ser Leu Pro Thr Ser Glu Tyr Asn Lys
            100                 105                 110
Ile Asn Ala Ile Ser Gln Ser Leu Gln Asn Thr Leu Glu Asn Lys Asn
        115                 120                 125
Asn Asp Leu Lys Ile Glu Asn Asp Tyr Asp His Leu Leu Thr Gln Ala
    130                 135                 140
Ser Thr Ile Ile Asn Thr Leu Gln Ser Gln Cys Pro Gly Ile Asp Gly
145                 150                 155                 160
Gly Asn Gly Lys Pro Trp Gly Ile Asn Ala Ser Gly Asn Ala Cys Asn
                165                 170                 175
Ile Phe Gly Asn Thr Phe Asn Ala Ile Thr Ser Met Ile Asp Ser Ala
            180                 185                 190
Lys Lys Ala Ala Ala Asp Ala Arg Arg Thr Ala Pro Glu Ser Pro Asn
        195                 200                 205
Gln Pro Ser Ala Phe Asn Asn Ala Asp Phe Asn Lys Asn Leu Asn Gln
    210                 215                 220
Val Ser Ser Val Ile Asn Asp Thr Ile Ser Tyr Leu Lys Gly Asp Asn
225                 230                 235                 240
Leu Ala Thr Ile Tyr Asn Thr Leu Gln Lys Thr Pro Asp Ser Lys Gly
                245                 250                 255
Phe Gln Ser Leu Val Ser Arg Ser Ser Tyr Ser Tyr Ser Leu Asn Glu
            260                 265                 270
Thr Gln Tyr Ser Glu Phe Gln Thr Thr Thr Lys Glu Phe Gly His Asn
        275                 280                 285
Pro Phe Arg Ser Val Gly Leu Ile Asn Ser Gln Ser Asn Asn Gly Ala
    290                 295                 300
Met Asn Gly Val Gly Val Gln Leu Gly Tyr Lys Gln Phe Phe Gly Lys
305                 310                 315                 320
Asn Lys Phe Phe Gly Ile Arg Tyr Tyr Ala Phe Phe Asp Tyr Asn His
                325                 330                 335
Ala Tyr Ile Lys Ser Asn Phe Phe Asn Ser Ala Ser Asn Val Phe Thr
            340                 345                 350
Tyr Gly Ala Gly Ser Asp Leu Leu Leu Asn Phe Ile Asn Gly Gly Ser
        355                 360                 365
Asp Lys Asn Arg Lys Val Ser Phe Gly Ile Phe Gly Gly Ile Ala Leu
    370                 375                 380
Ala Gly Thr Thr Trp Leu Asn Ser Gln Phe Met Asn Leu Lys Thr Thr
385                 390                 395                 400
Asn Ser Ala Tyr Ser Ala Lys Ile Asn Asn Thr Asn Phe Gln Phe Leu
                405                 410                 415
Phe Asn Thr Gly Leu Arg Leu Gln Gly Ile His His Gly Val Glu Leu
            420                 425                 430
Gly Val Lys Ile Pro Thr Ile Asn Thr Asn Tyr Tyr Ser Phe Met Gly
        435                 440                 445
Ala Lys Leu Ala Tyr Arg Arg Leu Tyr Ser Val Tyr Phe Asn Tyr Val
    450                 455                 460
Leu Ala Tyr
465
```

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2338 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 146...2218
(D) OTHER INFORMATION:

(A) NAME/KEY: Signal Sequence
(B) LOCATION: 146...200
(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

```
ACTTAAAATT GTTTTTTTTT TTTTTCAAAA TATAAATTTT AAGCCAAAAA TAAGCATTTT     60
ATGGTAAAAT GGCGAACTTT CATAAACATG ACTATTATGG GAATGTCATG GGAATGTGAA    120
GAAAAATCTA TTAAAA GGA GAA AAC ATG AAA AAA TCC CTC TTA CTC TCT CTT    172
                              Met Lys Lys Ser Leu Leu Leu Ser Leu
                              -18         -15                 -10

TCT CTC ATC GCT TCC TTA TCA AGA GCT GAA GAT GAC GGA TTT TAT ACG     220
Ser Leu Ile Ala Ser Leu Ser Arg Ala Glu Asp Asp Gly Phe Tyr Thr
                -5                   1               5

AGT GTG GGC TAT CAG ATC GGT GAA GCG GTC CAA CAA GTG AAA AAC ACA     268
Ser Val Gly Tyr Gln Ile Gly Glu Ala Val Gln Gln Val Lys Asn Thr
        10                  15                  20

GGA GCA TTG CAA AAT CTT GCA GAC AGA TAC GAT AAC TTA AAC AAC CTT     316
Gly Ala Leu Gln Asn Leu Ala Asp Arg Tyr Asp Asn Leu Asn Asn Leu
    25                  30                  35

TTA AAC CAA TAC AAT TAT TTA AAT TCC TTA GTC AAT TTA GCC AGC ACG     364
Leu Asn Gln Tyr Asn Tyr Leu Asn Ser Leu Val Asn Leu Ala Ser Thr
40                  45                  50                  55

CCG AGC GCG ATC ACC GGT GCG ATT GAT AAT TTA AGC TCA AGC GCG ATT     412
Pro Ser Ala Ile Thr Gly Ala Ile Asp Asn Leu Ser Ser Ser Ala Ile
                60                  65                  70

AAC CTC ACT AGC GCC ACC ACC ACT TCC CCC GCC TAT CAA GCT GTG GCT     460
Asn Leu Thr Ser Ala Thr Thr Thr Ser Pro Ala Tyr Gln Ala Val Ala
            75                  80                  85

TTA GCG CTC AAT GCC GCT GTG GGC ATG TGG CAA GTC ATA GCC CTT TTT     508
Leu Ala Leu Asn Ala Ala Val Gly Met Trp Gln Val Ile Ala Leu Phe
```

```
        90                  95                  100

ATT GGC TGT GGC CCT GGC CCT ACC AAT AAT CAA AGC TAT CAA TCG TTT      556
Ile Gly Cys Gly Pro Gly Pro Thr Asn Asn Gln Ser Tyr Gln Ser Phe
    105                 110                 115

GGT AAC ACA CCA GCC CTT AAT GGG ACC ACC ACC ACT TGC AAT CAA GCA      604
Gly Asn Thr Pro Ala Leu Asn Gly Thr Thr Thr Thr Cys Asn Gln Ala
120                 125                 130                 135

TAT GGG ACA GGC CCT AAT GGC ATC CTA TCT ATT GAT GAA TAC CAA AAA      652
Tyr Gly Thr Gly Pro Asn Gly Ile Leu Ser Ile Asp Glu Tyr Gln Lys
                140                 145                 150

CTC AAC CAA GCT TAT CAG ATC ATC CAA ACC GCT TTA AAC CAA AAT CAA      700
Leu Asn Gln Ala Tyr Gln Ile Ile Gln Thr Ala Leu Asn Gln Asn Gln
            155                 160                 165

GGG GGT GGG ATG CCT GCC TTG AAT GAC ACC ACC AAA ACA GGG GTA GTC      748
Gly Gly Gly Met Pro Ala Leu Asn Asp Thr Thr Lys Thr Gly Val Val
        170                 175                 180

AAC ATA CAA CAA ACC AAT TAT AGG ACC ACC ACA CAA AAC AAT ATC ATA      796
Asn Ile Gln Gln Thr Asn Tyr Arg Thr Thr Thr Gln Asn Asn Ile Ile
    185                 190                 195

GAG CAT TAT TAT ACA GAG AAT GGG AAA GAG ATC CCA GTC TCT TAT TCA      844
Glu His Tyr Tyr Thr Glu Asn Gly Lys Glu Ile Pro Val Ser Tyr Ser
200                 205                 210                 215

GGC GGA TCA TCA TTC TCG CCT ACA ATA CAA TTG ACA TAC CAT AAT AAC      892
Gly Gly Ser Ser Phe Ser Pro Thr Ile Gln Leu Thr Tyr His Asn Asn
                220                 225                 230

GCT GAA AAC CTT TTG CAA CAA GCC GCC ACT ATC ATG CAA GTC CTT ATT      940
Ala Glu Asn Leu Leu Gln Gln Ala Ala Thr Ile Met Gln Val Leu Ile
            235                 240                 245

ACT CAA AAG CCG CAT GTG CAA ACG AGC AAT GGC GGT AAA GCG TGG GGG      988
Thr Gln Lys Pro His Val Gln Thr Ser Asn Gly Gly Lys Ala Trp Gly
        250                 255                 260

TTG AGT TCT ACG CCT GGG AAT GTG ATG GAT ATT TTT GGT CCT TCT TTT     1036
Leu Ser Ser Thr Pro Gly Asn Val Met Asp Ile Phe Gly Pro Ser Phe
    265                 270                 275

AAC GCT ATT AAT GAG ATG ATT AAA AAC GCT CAA ACA GCC CTA GCA AAA     1084
Asn Ala Ile Asn Glu Met Ile Lys Asn Ala Gln Thr Ala Leu Ala Lys
280                 285                 290                 295

ACC CAA CAG CTT AAC GCT AAT GAA AAC GCC CAA ATC ACG CAA CCC AAC     1132
Thr Gln Gln Leu Asn Ala Asn Glu Asn Ala Gln Ile Thr Gln Pro Asn
                300                 305                 310

AAT TTC AAC CCC TAC ACC TCT AAA GAC AAA GGG TTC GCT CAA GAA ATG     1180
Asn Phe Asn Pro Tyr Thr Ser Lys Asp Lys Gly Phe Ala Gln Glu Met
            315                 320                 325
```

```
CTC AAT AGA GCT GAA GCT CAA GCA GAG ATT TTA AAT TTA GCT AAG CAA    1228
Leu Asn Arg Ala Glu Ala Gln Ala Glu Ile Leu Asn Leu Ala Lys Gln
        330                 335                 340

GTA GCG AAC AAT TTC CAC AGC ATT CAA GGG CCT ATT CAA GGG GAT TTA    1276
Val Ala Asn Asn Phe His Ser Ile Gln Gly Pro Ile Gln Gly Asp Leu
    345                 350                 355

GAA GAA TGT AAA GCA GGA TCG GCT GGC GTG ATC ACT AAT AAC ACT TGG    1324
Glu Glu Cys Lys Ala Gly Ser Ala Gly Val Ile Thr Asn Asn Thr Trp
360                 365                 370                 375

GGT TCA GGT TGC GCG TTT GTG AAA GAA ACT TTA AAC TCT TTA GAG CAA    1372
Gly Ser Gly Cys Ala Phe Val Lys Glu Thr Leu Asn Ser Leu Glu Gln
                380                 385                 390

CAC ACC GCT TAT TAC GGC AAC CAG GTC AAT CAG GAT AGG GCT TTG GCT    1420
His Thr Ala Tyr Tyr Gly Asn Gln Val Asn Gln Asp Arg Ala Leu Ala
            395                 400                 405

CAA ACC ATT TTG AAT TTT AAA GAA GCC CTT AAC ACC CTG AAT AAA GAC    1468
Gln Thr Ile Leu Asn Phe Lys Glu Ala Leu Asn Thr Leu Asn Lys Asp
        410                 415                 420

TCA AAA GCG ATC AAT AGC GGT ATC TCC AAC TTG CCT AAC GCT AAA TCT    1516
Ser Lys Ala Ile Asn Ser Gly Ile Ser Asn Leu Pro Asn Ala Lys Ser
    425                 430                 435

CTT CAA AAC ATG ACG CAT GCC ACT CAA AAC CCT AAT TCC CCA GAA GGT    1564
Leu Gln Asn Met Thr His Ala Thr Gln Asn Pro Asn Ser Pro Glu Gly
440                 445                 450                 455

CTG CTC ACT TAT TCT TTG GAT TCA AGC AAA TAC AAC CAG CTC CAA ACC    1612
Leu Leu Thr Tyr Ser Leu Asp Ser Ser Lys Tyr Asn Gln Leu Gln Thr
                460                 465                 470

ATC GCG CAA GAA TTG GGC AAA AAC CCT TTC AGG CGC TTT GGC GTG ATT    1660
Ile Ala Gln Glu Leu Gly Lys Asn Pro Phe Arg Arg Phe Gly Val Ile
            475                 480                 485

GAC TTT CAA AAC AAC AAC GGC GCA ATG AAC GGG ATC GGC GTG CAA GTG    1708
Asp Phe Gln Asn Asn Asn Gly Ala Met Asn Gly Ile Gly Val Gln Val
        490                 495                 500

GGT TAT AAA CAA TTC TTT GGT AAA AAA AGG AAT TGG GGG TTA AGG TAT    1756
Gly Tyr Lys Gln Phe Phe Gly Lys Lys Arg Asn Trp Gly Leu Arg Tyr
    505                 510                 515

TAT GGT TTC TTT GAT TAT AAC CAT GCT TAT ATC AAA TCT AAT TTT TTC    1804
Tyr Gly Phe Phe Asp Tyr Asn His Ala Tyr Ile Lys Ser Asn Phe Phe
520                 525                 530                 535

AAC TCC GCT TCT GAT GTG TGG ACT TAT GGG GTG GGT ATG GAC GCT CTC    1852
Asn Ser Ala Ser Asp Val Trp Thr Tyr Gly Val Gly Met Asp Ala Leu
                540                 545                 550

TAT AAC TTC ATC AAC GAT AAA AAC ACC AAC TTT TTA GGC AAG AAC AAC    1900
```

```
Tyr Asn Phe Ile Asn Asp Lys Asn Thr Asn Phe Leu Gly Lys Asn Asn
            555                 560                 565

AAG CTT TCA GTA GGG CTT TTT GGA GGC TTT GCG TTA GCC GGG ACT TCG    1948
Lys Leu Ser Val Gly Leu Phe Gly Gly Phe Ala Leu Ala Gly Thr Ser
        570                 575                 580

TGG CTT AAT TCC CAA CAA GTG AAT TTG ACC ATG ATG AAT GGC ATT TAT    1996
Trp Leu Asn Ser Gln Gln Val Asn Leu Thr Met Met Asn Gly Ile Tyr
    585                 590                 595

AAC GCT AAT GTC AGC ACT TCT AAC TTC CAA TTT TTG TTT GAT TTA GGC    2044
Asn Ala Asn Val Ser Thr Ser Asn Phe Gln Phe Leu Phe Asp Leu Gly
600                 605                 610                 615

TTG AGA ATG AAC CTC GCT AGG CCT AAG AAA AAA GAC AGC GAT CAT GCC    2092
Leu Arg Met Asn Leu Ala Arg Pro Lys Lys Lys Asp Ser Asp His Ala
                620                 625                 630

GCT CAG CAT GGC ATT GAA CTA GGT TTT AAG ATC CCC ACG ATC AAC ACC    2140
Ala Gln His Gly Ile Glu Leu Gly Phe Lys Ile Pro Thr Ile Asn Thr
            635                 640                 645

AAC TAT TAT TCT TTC ATG GGC GCT AAA CTA GAA TAC AGA AGG ATG TAT    2188
Asn Tyr Tyr Ser Phe Met Gly Ala Lys Leu Glu Tyr Arg Arg Met Tyr
        650                 655                 660

AGC CTT TTT CTC AAT TAT GTG TTT GCT TAC TAAAAACTCT CTTTAAAAAA GGG  2241
Ser Leu Phe Leu Asn Tyr Val Phe Ala Tyr
    665                 670

GTTTGTTTAA AAACGCTTAA AAGCATTTTT AAAATTAAGC AGTAAAGAGC CTAGATAATC   2301
TCTTGCAACC GCTCTCAAGC GATAAAATTA AAGTGAT                             2338
```

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 691 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal
(ix) FEATURE:

(A) NAME/KEY: Signal Sequence
(B) LOCATION: 1...18
(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

```
Met Lys Lys Ser Leu Leu Leu Ser Leu Ser Leu Ile Ala Ser Leu Ser
-18          -15                 -10                 -5
Arg Ala Glu Asp Asp Gly Phe Tyr Thr Ser Val Gly Tyr Gln Ile Gly
```

```
          1               5                  10
Glu Ala Val Gln Gln Val Lys Asn Thr Gly Ala Leu Gln Asn Leu Ala
15                  20                  25                  30
Asp Arg Tyr Asp Asn Leu Asn Asn Leu Leu Asn Gln Tyr Asn Tyr Leu
                35                  40                  45
Asn Ser Leu Val Asn Leu Ala Ser Thr Pro Ser Ala Ile Thr Gly Ala
            50                  55                  60
Ile Asp Asn Leu Ser Ser Ser Ala Ile Asn Leu Thr Ser Ala Thr Thr
        65                  70                  75
Thr Ser Pro Ala Tyr Gln Ala Val Ala Leu Ala Leu Asn Ala Ala Val
    80                  85                  90
Gly Met Trp Gln Val Ile Ala Leu Phe Ile Gly Cys Gly Pro Gly Pro
95                  100                 105                 110
Thr Asn Asn Gln Ser Tyr Gln Ser Phe Gly Asn Thr Pro Ala Leu Asn
                115                 120                 125
Gly Thr Thr Thr Thr Cys Asn Gln Ala Tyr Gly Thr Gly Pro Asn Gly
            130                 135                 140
Ile Leu Ser Ile Asp Glu Tyr Gln Lys Leu Asn Gln Ala Tyr Gln Ile
        145                 150                 155
Ile Gln Thr Ala Leu Asn Gln Asn Gln Gly Gly Gly Met Pro Ala Leu

    160                 165                 170
Asn Asp Thr Thr Lys Thr Gly Val Val Asn Ile Gln Gln Thr Asn Tyr
175                 180                 185                 190
Arg Thr Thr Thr Gln Asn Asn Ile Ile Glu His Tyr Tyr Thr Glu Asn
                195                 200                 205
Gly Lys Glu Ile Pro Val Ser Tyr Ser Gly Gly Ser Ser Phe Ser Pro
            210                 215                 220
Thr Ile Gln Leu Thr Tyr His Asn Asn Ala Glu Asn Leu Leu Gln Gln
        225                 230                 235
Ala Ala Thr Ile Met Gln Val Leu Ile Thr Gln Lys Pro His Val Gln
    240                 245                 250
Thr Ser Asn Gly Gly Lys Ala Trp Gly Leu Ser Ser Thr Pro Gly Asn
255                 260                 265                 270
Val Met Asp Ile Phe Gly Pro Ser Phe Asn Ala Ile Asn Glu Met Ile
                275                 280                 285
Lys Asn Ala Gln Thr Ala Leu Ala Lys Thr Gln Gln Leu Asn Ala Asn
            290                 295                 300
Glu Asn Ala Gln Ile Thr Gln Pro Asn Asn Phe Asn Pro Tyr Thr Ser
        305                 310                 315
Lys Asp Lys Gly Phe Ala Gln Glu Met Leu Asn Arg Ala Glu Ala Gln
    320                 325                 330
Ala Glu Ile Leu Asn Leu Ala Lys Gln Val Ala Asn Asn Phe His Ser
335                 340                 345                 350
Ile Gln Gly Pro Ile Gln Gly Asp Leu Glu Glu Cys Lys Ala Gly Ser
                355                 360                 365
Ala Gly Val Ile Thr Asn Asn Thr Trp Gly Ser Gly Cys Ala Phe Val
            370                 375                 380
Lys Glu Thr Leu Asn Ser Leu Glu Gln His Thr Ala Tyr Tyr Gly Asn
        385                 390                 395
Gln Val Asn Gln Asp Arg Ala Leu Ala Gln Thr Ile Leu Asn Phe Lys
    400                 405                 410
Glu Ala Leu Asn Thr Leu Asn Lys Asp Ser Lys Ala Ile Asn Ser Gly
415                 420                 425                 430
Ile Ser Asn Leu Pro Asn Ala Lys Ser Leu Gln Asn Met Thr His Ala
                435                 440                 445
Thr Gln Asn Pro Asn Ser Pro Glu Gly Leu Leu Thr Tyr Ser Leu Asp
```

```
            450                 455                 460
Ser Ser Lys Tyr Asn Gln Leu Gln Thr Ile Ala Gln Glu Leu Gly Lys
        465                 470                 475
Asn Pro Phe Arg Arg Phe Gly Val Ile Asp Phe Gln Asn Asn Asn Gly
    480                 485                 490
Ala Met Asn Gly Ile Gly Val Gln Val Gly Tyr Lys Gln Phe Phe Gly
495                 500                 505                 510
Lys Lys Arg Asn Trp Gly Leu Arg Tyr Tyr Gly Phe Phe Asp Tyr Asn
                515                 520                 525
His Ala Tyr Ile Lys Ser Asn Phe Phe Asn Ser Ala Ser Asp Val Trp
            530                 535                 540
Thr Tyr Gly Val Gly Met Asp Ala Leu Tyr Asn Phe Ile Asn Asp Lys
        545                 550                 555
Asn Thr Asn Phe Leu Gly Lys Asn Asn Lys Leu Ser Val Gly Leu Phe
    560                 565                 570
Gly Gly Phe Ala Leu Ala Gly Thr Ser Trp Leu Asn Ser Gln Gln Val
575                 580                 585                 590
Asn Leu Thr Met Met Asn Gly Ile Tyr Asn Ala Asn Val Ser Thr Ser
                595                 600                 605
Asn Phe Gln Phe Leu Phe Asp Leu Gly Leu Arg Met Asn Leu Ala Arg
            610                 615                 620
Pro Lys Lys Lys Asp Ser Asp His Ala Ala Gln His Gly Ile Glu Leu
        625                 630                 635
Gly Phe Lys Ile Pro Thr Ile Asn Thr Asn Tyr Tyr Ser Phe Met Gly
    640                 645                 650
Ala Lys Leu Glu Tyr Arg Arg Met Tyr Ser Leu Phe Leu Asn Tyr Val
655                 660                 665                 670
Phe Ala Tyr
```

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

TCAAGGAGAA AACATGAAAA AAACCC 26

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GAAGACGACG GCTTTTACAC AAGCGT 26

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

AAAGCTTAGT AAGCGAACAC ATAA 24

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

AAGGAGAAAA AACATGAAAA AACACATCC 29

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GAAGACGACG GCTTTTACAC AAGCG 25

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

AACATTAGTA AGCGAACACA TAGTTC 26

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

AAGGAGAAAA AACATGAAAA AACACATCC 29

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

GAAGACGACG GCTTTTACAC AAGCGT 26

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

AAAAGCTTAG TAAGCGAACA CAT 23

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

AAGGAGAAAA CATGAAGAAA AAATTT 26

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

GAAGACAACG GCTTTTTTGT GAGTG 25

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

AGCTTTTAGT AAGCAAACAC ATAGT 25

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

AAGGATATTT ATGAAAAAAA CCCTT 25

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

GAAGACAACG GCTTTTTTAT CAGCG 25

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

GATATTAGTA AGCAAACACA TAATTC 26

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

AAGGAGAAAA CATGAAAAAA TCCCTCT 27

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

GAAGATGACG GATTTTATAC GAGTGT 26

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

TTTTAGTAAG CAAACACATA ATTGAG 26

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

AAGGAACATC TTATGAAAAA AACG 24

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

GAAGACAACG GCGTTTTTTT AAGCG 25

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

GGTTTTTAAT AGGCAAACAC ATAAT 25

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

AAGGAACATT TTATGAAAAA GACAAT 26

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

GAAGACAACG GCTTTTTTGT GAGCG 25

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

TCACTCAGTA AGCGAACACA TAA 23

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

AAGGAACATT TTATGAAAAA GACAA 25

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

GAAGACAACG GCTTTTTTGT GAGCG 25

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

TTTTAATAAG CGAACACATA AAAGAG 26

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

AAGGAACATT TTATGAAAAA AACGAT 26

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

GAAAATGACG GCGTTTTTAT GAGCG 25

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

ATATCAATAG GCCAAAACAT AATTGA 26

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

AAGGAGAAAA CATGAAAAAA TCCCTC 26

(2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

GAAGATGACG GATTTTATAC GAGTGT 26

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

TTTTAGTAAG CAAACACATA ATTGAG 26

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

CGCGGATCCG AATCCAATTT AATCCAAAAA GG 32

(2) INFORMATION FOR SEQ ID NO:57:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

CCGCTCGAGT TAAGTAAGCG AACACATATT CAA 33

(2) INFORMATION FOR SEQ ID NO:58

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58

Glu Asp Asp Gly Phe Tyr Thr Ser Val Gly Tyr Gln Ile Gly Glu Ala
1 5 10 15
Ala Gln Met Val
20

(2) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

CTGAATTCGA TTTCAAGGAG AAAACATGAA A 31

(2) INFORMATION FOR SEQ ID NO:60:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

CCGCTCGAGT TAGTAAGCGA ACACATAATT 30

(2) INFORMATION FOR SEQ ID NO:61:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

CGCGGATCCG AATCCAATTT AATCCAAAAA GG 32

(2) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

CCGCTCGAGT TAGTAAGCGA ACACATAGTT CAA 33

(2) INFORMATION FOR SEQ ID NO:63:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

CGCGGATCCG AAGTTTCTTT GTATCAAAG 29

(2) INFORMATION FOR SEQ ID NO:64:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

CCGCTCGAGT TAGTAAGCAA ACACATAATT GTG 33

(2) INFORMATION FOR SEQ ID NO:65:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1149 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 106...1002
(D) OTHER INFORMATION:

(A) NAME/KEY: Signal Sequence
(B) LOCATION: 106...166
(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

```
TTACTCTTTA ATGTGAGTTT TCTGTGTCAT GATAGCTGAT TTTGTTTTAA ATTTGCTATA     60
ATGTGAATTT AATGATGAAA ATTAGTTTAG AGTGGAGAAC ACACA ATG AAA AAA AAT     117
                                                    Met Lys Lys Asn
                                                    -20
ATC TTA AAT TTA GCG TTA GTG GGT GCG TTG AGC ACG TCG TTT TTG ATG       165
Ile Leu Asn Leu Ala Leu Val Gly Ala Leu Ser Thr Ser Phe Leu Met
    -15                 -10                 -5
GCT AAG CCG GCT CAT AAC GCA AAT AAC GCT ACG CAT AAC ACG AAA AAA       213
Ala Lys Pro Ala His Asn Ala Asn Asn Ala Thr His Asn Thr Lys Lys
 1               5                  10                  15
ACG ACT GAT TCT TCA GCA GGC GTG TTA GCG ACA GTG GAT GGC AGA CCT       261
Thr Thr Asp Ser Ser Ala Gly Val Leu Ala Thr Val Asp Gly Arg Pro
            20                  25                  30
ATC ACT AAA AGC GAT TTT GAC ATG ATT AAG CAA CGA AAT CCT AAT TTT       309
Ile Thr Lys Ser Asp Phe Asp Met Ile Lys Gln Arg Asn Pro Asn Phe
        35                  40                  45
GAT TTT GAC AAG CTT AAA GAG AAA GAA AAA GAA GCC TTG ATT GAT CAA       357
Asp Phe Asp Lys Leu Lys Glu Lys Glu Lys Glu Ala Leu Ile Asp Gln
    50                  55                  60
GCT ATT CGC ACC GCC CTT GTA GAA AAT GAA GCT AAA ACC GAG AAA TTG       405
Ala Ile Arg Thr Ala Leu Val Glu Asn Glu Ala Lys Thr Glu Lys Leu
65                  70                  75                  80
GAC AGC ACT CCA GAA TTT AAA GCG ATG ATG GAA GCG GTT AAA AAA CAG       453
Asp Ser Thr Pro Glu Phe Lys Ala Met Met Glu Ala Val Lys Lys Gln
                85                  90                  95
GCT TTA GTG GAA TTT TGG GCT AAA AAA CAG GCT GAA GAA GTG AAA AAA       501
Ala Leu Val Glu Phe Trp Ala Lys Lys Gln Ala Glu Glu Val Lys Lys
            100                 105                 110
GTC CAA ATC CCA GAA AAA GAA ATG CAA GAT TTT TAC AAC GCT AAC AAA       549
Val Gln Ile Pro Glu Lys Glu Met Gln Asp Phe Tyr Asn Ala Asn Lys
        115                 120                 125
GAT CAG CTT TTT GTC AAG CAA GAA GCC CAT GCT AGG CAT ATT TTA GTG       597
Asp Gln Leu Phe Val Lys Gln Glu Ala His Ala Arg His Ile Leu Val
    130                 135                 140
AAA ACC GAA GAT GAG GCT AAA CGG ATT ATT TCT GAG ATT GAC AAA CAG       645
Lys Thr Glu Asp Glu Ala Lys Arg Ile Ile Ser Glu Ile Asp Lys Gln
145                 150                 155                 160
CCA AAG GCT AAA AAA GAA GCT AAA TTC ATT GAG TTA GCC AAT CGG GAT       693
Pro Lys Ala Lys Lys Glu Ala Lys Phe Ile Glu Leu Ala Asn Arg Asp
                165                 170                 175
```

```
ACG ATT GAT CCT AAC AGC AAG AAC GCG CAA AAT GGC GGT GAT TTG GGG      741
Thr Ile Asp Pro Asn Ser Lys Asn Ala Gln Asn Gly Gly Asp Leu Gly
            180                 185                 190
AAA TTC CAA AAG AAC CAA ATG GCT CCG GAT TTT TCT AAA GCC GCT TTC      789
Lys Phe Gln Lys Asn Gln Met Ala Pro Asp Phe Ser Lys Ala Ala Phe
        195                 200                 205
GCT TTA ACT CCT GGG GAT TAC ACT AAA ACC CCT GTT AAA ACA GAG TTT      837
Ala Leu Thr Pro Gly Asp Tyr Thr Lys Thr Pro Val Lys Thr Glu Phe
    210                 215                 220
GGT TAT CAT ATT ATC TAT TTG ATT TCT AAA GAT AGC CCT GTA ACT TAT      885
Gly Tyr His Ile Ile Tyr Leu Ile Ser Lys Asp Ser Pro Val Thr Tyr
225                 230                 235                 240
ACT TAT GAA CAG GCT AAA CCT ACC ATT AAG GGG ATG TTA CAA GAA AAG      933
Thr Tyr Glu Gln Ala Lys Pro Thr Ile Lys Gly Met Leu Gln Glu Lys
                245                 250                 255
CTT TTC CAA GAA CGC ATG AAT CAA CGC ATT GAG GAA CTA AGA AAG CAC      981
Leu Phe Gln Glu Arg Met Asn Gln Arg Ile Glu Glu Leu Arg Lys His
            260                 265                 270
GCT AAA ATT GTT ATC AAC AAG TAATTGATGA GGTGTTATCA TGTTAGTTAA AGGC  1036
Ala Lys Ile Val Ile Asn Lys
        275
AATGAAATTT TATTGAAAGC CCATAAAGAA GGTTATGGGG TGGGGGCGTT TAATTTCGTG  1096
AATTTTGAAA TGCTAAACGC TATTTTTGAA GCAGGAAATG AGGAAAATTC CCC         1149
```

(2) INFORMATION FOR SEQ ID NO:66:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 299 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal
(ix) FEATURE:

(A) NAME/KEY: Signal Sequence
(B) LOCATION: 1...20
(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

```
Met Lys Lys Asn Ile Leu Asn Leu Ala Leu Val Gly Ala Leu Ser Thr
-20                 -15                 -10                 -5
Ser Phe Leu Met Ala Lys Pro Ala His Asn Ala Asn Asn Ala Thr His
                 1                   5                  10
Asn Thr Lys Lys Thr Thr Asp Ser Ser Ala Gly Val Leu Ala Thr Val
        15                  20                  25
Asp Gly Arg Pro Ile Thr Lys Ser Asp Phe Asp Met Ile Lys Gln Arg
    30                  35                  40
Asn Pro Asn Phe Asp Phe Asp Lys Leu Lys Glu Lys Glu Lys Glu Ala
45                  50                  55                  60
Leu Ile Asp Gln Ala Ile Arg Thr Ala Leu Val Glu Asn Glu Ala Lys
                65                  70                  75
Thr Glu Lys Leu Asp Ser Thr Pro Glu Phe Lys Ala Met Met Glu Ala
```

```
            80                  85                  90
Val Lys Lys Gln Ala Leu Val Glu Phe Trp Ala Lys Lys Gln Ala Glu
        95                  100                 105
Glu Val Lys Lys Val Gln Ile Pro Glu Lys Glu Met Gln Asp Phe Tyr
    110                 115                 120
Asn Ala Asn Lys Asp Gln Leu Phe Val Lys Gln Glu Ala His Ala Arg
125                 130                 135                 140
His Ile Leu Val Lys Thr Glu Asp Glu Ala Lys Arg Ile Ile Ser Glu
                145                 150                 155
Ile Asp Lys Gln Pro Lys Ala Lys Lys Glu Ala Lys Phe Ile Glu Leu
            160                 165                 170
Ala Asn Arg Asp Thr Ile Asp Pro Asn Ser Lys Asn Ala Gln Asn Gly
        175                 180                 185
Gly Asp Leu Gly Lys Phe Gln Lys Asn Gln Met Ala Pro Asp Phe Ser
    190                 195                 200
Lys Ala Ala Phe Ala Leu Thr Pro Gly Asp Tyr Thr Lys Thr Pro Val
205                 210                 215                 220
Lys Thr Glu Phe Gly Tyr His Ile Ile Tyr Leu Ile Ser Lys Asp Ser
                225                 230                 235
Pro Val Thr Tyr Thr Tyr Glu Gln Ala Lys Pro Thr Ile Lys Gly Met
            240                 245                 250
Leu Gln Glu Lys Leu Phe Gln Glu Arg Met Asn Gln Arg Ile Glu Glu
        255                 260                 265
Leu Arg Lys His Ala Lys Ile Val Ile Asn Lys
    270                 275
```

(2) INFORMATION FOR SEQ ID NO:67:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1448 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 118...1314
(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

```
CTCTTGAATG GCGATAAGAC AAAAATGTCT TAAATTTTGT GGTAGCATTT AGGAATACTT     60
AGGATTTTGT TTAGTATAAT TCTAAAATCC ATTTCAAAAA ATTAAGGAGA AATACAA ATG   120
                                                                 Met
                                                                  1
GCA AAA GAA AAG TTT AAC AGA ACT AAG CCG CAT GTT AAT ATT GGA ACC      168
Ala Lys Glu Lys Phe Asn Arg Thr Lys Pro His Val Asn Ile Gly Thr
          5                   10                  15
ATT GGG CAT GTA GAC CAT GGT AAA ACG ACT TTG AGT GCA GCG ATT TCA      216
Ile Gly His Val Asp His Gly Lys Thr Thr Leu Ser Ala Ala Ile Ser
      20                  25                  30
GCG GTG CTT TCT TTG AAA GGT CTT GCA GAA ATG AAA GAC TAT GAT AAT      264
Ala Val Leu Ser Leu Lys Gly Leu Ala Glu Met Lys Asp Tyr Asp Asn
```

```
    35                  40                  45
ATT GAT AAC GCC CCT GAA GAA AAA GAA AGA GGG ATC ACT ATC GCT ACT      312
Ile Asp Asn Ala Pro Glu Glu Lys Glu Arg Gly Ile Thr Ile Ala Thr
50                  55                  60                  65
TCT CAC ATT GAA TAT GAG ACT GAA AAC AGA CAC TAT GCG CAT GTG GAT      360
Ser His Ile Glu Tyr Glu Thr Glu Asn Arg His Tyr Ala His Val Asp
                70                  75                  80
TGC CCA GGA CAC GCT GAC TAT GTA AAA AAC ATG ATC ACC GGT GCG GCG      408
Cys Pro Gly His Ala Asp Tyr Val Lys Asn Met Ile Thr Gly Ala Ala
            85                  90                  95
CAA ATG GAC GGA GCG ATT TTG GTT GTT TCT GCA GCT GAT GGC CCT ATG      456
Gln Met Asp Gly Ala Ile Leu Val Val Ser Ala Ala Asp Gly Pro Met
        100                 105                 110
CCT CAA ACT AGG GAG CAT ATC TTA TTG TCT CGT CAA GTA GGC GTG CCT      504
Pro Gln Thr Arg Glu His Ile Leu Leu Ser Arg Gln Val Gly Val Pro
    115                 120                 125
CAC ATC GTT GTT TTC TTA AAC AAA CAA GAC ATG GTA GAT GAC CAA GAA      552
His Ile Val Val Phe Leu Asn Lys Gln Asp Met Val Asp Asp Gln Glu
130                 135                 140                 145
TTG TTA GAA CTT GTA GAA ATG GAA GTG CGC GAA TTG TTG AGC GCG TAT      600
Leu Leu Glu Leu Val Glu Met Glu Val Arg Glu Leu Leu Ser Ala Tyr
                150                 155                 160
GAA TTT CCT GGC GAT GAC ACT CCT ATC GTA GCG GGT TCA GCT TTA AGA      648
Glu Phe Pro Gly Asp Asp Thr Pro Ile Val Ala Gly Ser Ala Leu Arg
            165                 170                 175
GCT TTA GAA GAA GCA AAG GCT GGT AAT GTG GGT GAA TGG GGT GAA AAA      696
Ala Leu Glu Glu Ala Lys Ala Gly Asn Val Gly Glu Trp Gly Glu Lys
        180                 185                 190
GTG CTT AAA CTT ATG GCT GAA GTG GAT GCC TAT ATC CCT ACT CCA GAA      744
Val Leu Lys Leu Met Ala Glu Val Asp Ala Tyr Ile Pro Thr Pro Glu
    195                 200                 205
AGA GAC ACT GAA AAA ACT TTC TTG ATG CCG GTT GAA GAT GTG TTC TCT      792
Arg Asp Thr Glu Lys Thr Phe Leu Met Pro Val Glu Asp Val Phe Ser
210                 215                 220                 225
ATT GCG GGT AGA GGG ACT GTG GTT ACA GGT AGG ATT GAA AGA GGC GTG      840
Ile Ala Gly Arg Gly Thr Val Val Thr Gly Arg Ile Glu Arg Gly Val
                230                 235                 240
GTG AAA GTA GGC GAT GAA GTG GAA ATC GTT GGT ATC AGA CCT ACA CAA      888
Val Lys Val Gly Asp Glu Val Glu Ile Val Gly Ile Arg Pro Thr Gln
            245                 250                 255
AAA ACG ACT GTA ACC GGT GTA GAA ATG TTT AGG AAA GAG TTG GAA AAA      936
Lys Thr Thr Val Thr Gly Val Glu Met Phe Arg Lys Glu Leu Glu Lys
        260                 265                 270
GGT GAA GCC GGC GAT AAT GTG GGC GTG CTT TTG AGA GGA ACT AAA AAA      984
Gly Glu Ala Gly Asp Asn Val Gly Val Leu Leu Arg Gly Thr Lys Lys
    275                 280                 285
GAA GAA GTG GAA CGC GGT ATG GTT CTA TGC AAA CCA GGT TCT ATC ACT     1032
Glu Glu Val Glu Arg Gly Met Val Leu Cys Lys Pro Gly Ser Ile Thr
290                 295                 300                 305
CCG CAC AAG AAA TTT GAG GGA GAA ATT TAT GTC CTT TCT AAA GAA GAA     1080
Pro His Lys Lys Phe Glu Gly Glu Ile Tyr Val Leu Ser Lys Glu Glu
                310                 315                 320
GGC GGG AGA CAC ACT CCA TTC TTC ACC AAT TAC CGC CCG CAA TTC TAT     1128
Gly Gly Arg His Thr Pro Phe Phe Thr Asn Tyr Arg Pro Gln Phe Tyr
            325                 330                 335
GTG CGC ACA ACT GAT GTG ACT GGC TCT ATC ACC CTT CCT GAA GGC GTA     1176
Val Arg Thr Thr Asp Val Thr Gly Ser Ile Thr Leu Pro Glu Gly Val
```

```
        340                 345                 350
GAA ATG GTT ATG CCT GGC GAT AAT GTG AAA ATC ACT GTA GAG TTG ATT    1224
Glu Met Val Met Pro Gly Asp Asn Val Lys Ile Thr Val Glu Leu Ile
    355                 360                 365
AGC CCT GTT GCG TTA GAG TTG GGA ACT AAA TTT GCG ATT CGT GAA GGC    1272
Ser Pro Val Ala Leu Glu Leu Gly Thr Lys Phe Ala Ile Arg Glu Gly
370                 375                 380                 385
GGT AGG ACC GTT GGT GCT GGT GTT GTG AGC AAT ATT ATT GAA TAATATTAG  1323
Gly Arg Thr Val Gly Ala Gly Val Val Ser Asn Ile Ile Glu
                390                 395
CAAAAAGAGA GTTACCATAA AGGGTCATTA TGAAAGTTAA AATAGGGTTG AAGTGTTCTG  1383
ATTGTGAAGA TATCAATTAC AGCACAACCA AGAACGCTAA AACTAACACT GAAAAACTGG  1443
AGCTT                                                              1448
```

(2) INFORMATION FOR SEQ ID NO:68:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 399 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

```
Met Ala Lys Glu Lys Phe Asn Arg Thr Lys Pro His Val Asn Ile Gly
 1               5                  10                  15
Thr Ile Gly His Val Asp His Gly Lys Thr Thr Leu Ser Ala Ala Ile
            20                  25                  30
Ser Ala Val Leu Ser Leu Lys Gly Leu Ala Glu Met Lys Asp Tyr Asp
        35                  40                  45
Asn Ile Asp Asn Ala Pro Glu Glu Lys Glu Arg Gly Ile Thr Ile Ala
    50                  55                  60
Thr Ser His Ile Glu Tyr Glu Thr Glu Asn Arg His Tyr Ala His Val
65                  70                  75                  80
Asp Cys Pro Gly His Ala Asp Tyr Val Lys Asn Met Ile Thr Gly Ala
                85                  90                  95
Ala Gln Met Asp Gly Ala Ile Leu Val Val Ser Ala Ala Asp Gly Pro
            100                 105                 110
Met Pro Gln Thr Arg Glu His Ile Leu Leu Ser Arg Gln Val Gly Val
        115                 120                 125
Pro His Ile Val Val Phe Leu Asn Lys Gln Asp Met Val Asp Asp Gln
    130                 135                 140
Glu Leu Leu Glu Leu Val Glu Met Glu Val Arg Glu Leu Leu Ser Ala
145                 150                 155                 160
Tyr Glu Phe Pro Gly Asp Asp Thr Pro Ile Val Ala Gly Ser Ala Leu
                165                 170                 175
Arg Ala Leu Glu Glu Ala Lys Ala Gly Asn Val Gly Glu Trp Gly Glu
            180                 185                 190
Lys Val Leu Lys Leu Met Ala Glu Val Asp Ala Tyr Ile Pro Thr Pro
        195                 200                 205
Glu Arg Asp Thr Glu Lys Thr Phe Leu Met Pro Val Glu Asp Val Phe
    210                 215                 220
Ser Ile Ala Gly Arg Gly Thr Val Val Thr Gly Arg Ile Glu Arg Gly
225                 230                 235                 240
```

```
Val Val Lys Val Gly Asp Glu Val Glu Ile Val Gly Ile Arg Pro Thr
                245                 250                 255
Gln Lys Thr Thr Val Thr Gly Val Glu Met Phe Arg Lys Glu Leu Glu
            260                 265                 270
Lys Gly Glu Ala Gly Asp Asn Val Gly Val Leu Leu Arg Gly Thr Lys
        275                 280                 285
Lys Glu Glu Val Glu Arg Gly Met Val Leu Cys Lys Pro Gly Ser Ile
    290                 295                 300
Thr Pro His Lys Lys Phe Glu Gly Glu Ile Tyr Val Leu Ser Lys Glu
305                 310                 315                 320
Glu Gly Gly Arg His Thr Pro Phe Phe Thr Asn Tyr Arg Pro Gln Phe
                325                 330                 335
Tyr Val Arg Thr Thr Asp Val Thr Gly Ser Ile Thr Leu Pro Glu Gly
            340                 345                 350
Val Glu Met Val Met Pro Gly Asp Asn Val Lys Ile Thr Val Glu Leu
        355                 360                 365
Ile Ser Pro Val Ala Leu Glu Leu Gly Thr Lys Phe Ala Ile Arg Glu
    370                 375                 380
Gly Gly Arg Thr Val Gly Ala Gly Val Val Ser Asn Ile Ile Glu
385                 390                 395
```

(2) INFORMATION FOR SEQ ID NO:69:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

CGCGGATCCG AATGAAAAAA AATATCTTAA AT 32

(2) INFORMATION FOR SEQ ID NO:70:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

CCGCTCGAGT TACTTGTTGA TAACAATTTT 30

(2) INFORMATION FOR SEQ ID NO:71:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

CGCGGATCCG AATGGCAAAA GAAAAGTTTA AC 32

(2) INFORMATION FOR SEQ ID NO:72:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

CCGCTCGAGT TATTCAATAA TATTGCTCAC 30

(2) INFORMATION FOR SEQ ID NO:73:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

Met Lys Glu Lys Phe Asn Arg Thr Lys Pro His Val Asn Ile Gly Thr
1 5 10 15
Ile Gly His Val Asp His
20

(2) INFORMATION FOR SEQ ID NO:74:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

Ala His Asn Ala Asn Asn Ala Thr His Asn Thr Lys Lys
1 5 10

(2) INFORMATION FOR SEQ ID NO:75:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:

Lys Pro Ala His Asn Ala
1 5

(2) INFORMATION FOR SEQ ID NO:76:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

Ile Asp Lys Gln Pro Lys Ala Lys Lys
1 5

(2) INFORMATION FOR SEQ ID NO:77:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

Phe Trp Ala Lys Lys Gln Ala Glu
1 5

(2) INFORMATION FOR SEQ ID NO:78:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:

GTGGAGAACA CACAATGAAA AAAAATATC 29

(2) INFORMATION FOR SEQ ID NO:79:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:

GCTAATATTA TTCAATAATA TTGCTCACAA C 31

(2) INFORMATION FOR SEQ ID NO:80:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:80:

GGAGAAATAC AAATGGCAAA AGAAAAG 27

(2) INFORMATION FOR SEQ ID NO:81:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:

GCTAATATTA TTCAATAATA TTGCTCACAA C 31

(2) INFORMATION FOR SEQ ID NO:82:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

CATAACGCAA ATAACGCTAC GCAT 24

(2) INFORMATION FOR SEQ ID NO:83:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:

GGGAATTCAA AAAAACGAAA AAAACG 26

(2) INFORMATION FOR SEQ ID NO:84:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:

CCCCTCGAGT TAATAGGCAA ACAC 24

What is claimed is:

1. An isolated polynucleotide that encodes:

(i) a polypeptide comprising an amino acid sequence that is homologous to the amino acid sequence of a *Helicobacter* membrane-associated polypeptide, wherein said amino acid sequence of said *Helicobacter* membrane-associated polypeptide is selected from the group consisting of the amino acid sequences as shown:

-in SEQ ID NO:2, beginning with an amino acid in any one of positions -19 to 5, preferably in position -19 or position 1, and ending with an amino acid in position 689 (GHPO 386);

-in SEQ ID NO:4, beginning with an amino acid in any one of positions -20 to 5, preferably in position -20 or position 1, and ending with an amino acid in position 713 (GHPO 789);

-in SEQ ID NO:6, beginning with an amino acid in any one of positions -20 to 5, preferably in position -20 or position 1, and ending with an amino acid in position 725 (GHPO 1516);

-in SEQ ID NO:8, beginning with an amino acid in any one of positions -20 to 5, preferably in position -20 or position 1, and ending with an amino acid in position 691 (GHPO 1197);

-in SEQ ID NO:10, beginning with an amino acid in any one of positions -20 to 5, preferably in position -20 or position 1, and ending with an amino acid in position 652 (GHPO 1180);

-in SEQ ID NO:12, beginning with an amino acid in any one of positions -18 to 5, preferably in position -18 or position 1, and ending with an amino acid in position 673 (GHPO 896);

-in SEQ ID NO:14, beginning with an amino acid in any one of positions

-21 to 5, preferably in position -21 or position 1, and ending with an amino acid in position 619 (GHPO 711);

-in SEQ ID NO:16, beginning with an amino acid in any one of positions -17 to 5, preferably in position -17 or position 1, and ending with an amino acid in position 635 (GHPO 190);

-in SEQ ID NO:18, beginning with an amino acid in any one of positions -19 to 5, preferably in position -19 or position 1, and ending with an amino acid in position 626 (GHPO 185);

-in SEQ ID NO:20, beginning with an amino acid in any one of positions -16 to 5, preferably in position -16 or position 1, and ending with an amino acid in position 467 (GHPO 1417);

-in SEQ ID NO:22, beginning with an amino acid in any one of positions -18 to 5, preferably in position -18 or position 1, and ending with an amino acid in position 673 (GHPO 1414);

- in SEQ ID NO:66, beginning with an amino acid in any one of the positions from -20 to 5, preferably in position -20 or position 1, and ending with an amino acid in position 279 (GHPO 1360); and

- in SEQ ID NO:68, beginning with an amino acid in position 1 and ending with an amino acid in position 399 (GHPO 750); or

(ii) a derivative of the polypeptide.

2. An isolated polynucleotide that encodes:

(i) a polypeptide comprising an amino acid sequence that is homologous to an amino acid sequence selected from the group consisting of the amino acid sequences as shown:

-in SEQ ID NO:2, beginning with amino acid in position -19 and ending with an amino acid in position 689 (GHPO 386);

-in SEQ ID NO:4, beginning with an amino acid in position -20 and ending with an amino acid in position 713 (GHPO 789);

-in SEQ ID NO:6, beginning with an amino acid in position -20 and ending with an amino acid in position 725 (GHPO 1516);

-in SEQ ID NO:8, beginning with an amino acid in position -20 and ending with an amino acid in position 691 (GHPO 1197);

-in SEQ ID NO:10, beginning with an amino acid in position -20 and ending with an amino acid in position 652 (GHPO 1180);

-in SEQ ID NO:12, beginning with an amino acid in position -18 and ending with an amino acid in position 673 (GHPO 896);

-in SEQ ID NO:14, beginning with an amino acid in position -21 and ending with an amino acid in position 619 (GHPO 711);

-in SEQ ID NO:16, beginning with an amino acid in position -17 and ending with an amino acid in position 635 (GHPO 190);

-in SEQ ID NO:18, beginning with an amino acid in position -19 and ending with an amino acid in position 626 (GHPO 185);

-in SEQ ID NO:20, beginning with an amino acid in position -16 and ending with an amino acid in position 467 (GHPO 1417);

-in SEQ ID NO:22, beginning with an amino acid in position -18 and ending with an amino acid in position 673 (GHPO 1414);

- in SEQ ID NO:66, beginning with an amino acid in position -20 and ending with an amino acid in position 279 (GHPO 1360); and

- in SEQ ID NO:68, beginning with an amino acid in position 1 and ending with an amino acid in position 399 (GHPO 750); or

(ii) a derivative of the polypeptide.

3. The isolated polynucleotide of claim 1, which encodes the mature form of:

(i) a polypeptide comprising an amino acid sequence that is homologous to an amino acid sequence selected from the group consisting of the amino acid sequences as shown:

-in SEQ ID NO:2, beginning with an amino acid in any one of positions -19 to 5, preferably in position -19 or position 1, and ending with an amino acid in position 689 (GHPO 386);

-in SEQ ID NO:4, beginning with an amino acid in any one of positions -20 to 5, preferably in position -20 or position 1, and ending with an amino acid in position 713 (GHPO 789);

-in SEQ ID NO:6, beginning with an amino acid in any one of positions -20 to 5, preferably in position -20 or position 1, and ending with an amino acid in position 725 (GHPO 1516);

-in SEQ ID NO:8, beginning with an amino acid in any one of positions -20 to 5, preferably in position -20 or position 1, and ending with an amino acid in position 691 (GHPO 1197);

-in SEQ ID NO:10, beginning with an amino acid in any one of positions -20 to 5, preferably in position -20 or position 1, and ending with an amino acid in position 652 (GHPO 1180);

-in SEQ ID NO:12, beginning with an amino acid in any one of positions -18 to 5, preferably in position -18 or position 1, and ending with an amino acid in position 673 (GHPO 896);

-in SEQ ID NO:14, beginning with an amino acid in any one of positions -21 to 5, preferably in position -21 or position 1, and ending with an amino acid in position 619 (GHPO 711);

-in SEQ ID NO:16, beginning with an amino acid in any one of positions

-17 to 5, preferably in position -17 or position 1, and ending with an amino acid in position 635 (GHPO 190);

-in SEQ ID NO:18, beginning with an amino acid in any one of positions -19 to 5, preferably in position -19 or position 1, and ending with an amino acid in position 626 (GHPO 185);

-in SEQ ID NO:20, beginning with an amino acid in any one of positions -16 to 5, preferably in position -16 or position 1, and ending with an amino acid in position 467 (GHPO 1417);

-in SEQ ID NO:22, beginning with an amino acid in any one of positions -18 to 5, preferably in position -18 or position 1, and ending with an amino acid in position 673 (GHPO 1414);

- in SEQ ID NO:66, beginning with an amino acid in any one of positions
-20 to 5, preferably in position -20 or position 1, and ending with an amino acid in position 279 (GHPO 1360); and

- in SEQ ID NO:68, beginning with an amino acid in position 1 and ending with an amino acid in position 399 (GHPO 750); or

(ii) a derivative of the polypeptide.

4. The isolated polynucleotide of claim 1, 2, or 3, wherein the polynucleotide is a DNA molecule.

5. The isolated polynucleotide of claim 1, which is a DNA molecule that can be amplified and/or cloned by polymerase chain reaction from an *Helicobacter* genome, using either:

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:23, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:25 (unprocessed GHPO 386);

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:26, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:28 (unprocessed GHPO 789);

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:29, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:31 (unprocessed GHPO 1516);

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:32, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:34 (unprocessed GHPO 1197);

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:35, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:37 (unprocessed GHPO 1180);

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:38, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:40 (unprocessed GHPO 896);

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:41, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:43 (unprocessed GHPO 711);

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:44, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:46 (unprocessed GHPO 190);

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:47, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:49 (unprocessed GHPO 185);

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:50, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:52 (unprocessed GHPO 1417);

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:53, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:55 (unprocessed GHPO 1414);

- a 5' oligonucleotide primer comprising a sequence as shown in SEQ ID NO:78 and a 3' oligonucleotide primer comprising a sequence as shown in SEQ ID NO:79 (unprocessed GHPO 1360);

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:24, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:25 (mature GHPO 386);

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:27, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:28 (mature GHPO 789);

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:30, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:31 (mature GHPO 1516);

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:33, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:34 (mature GHPO 1197);

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:36, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:37 (mature GHPO 1180);

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:39, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:40 (mature GHPO 896);

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:42, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:43 (mature GHPO 711);

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:45, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:46 (mature GHPO 190);

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:48, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:49 (mature GHPO 185);

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:51, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:52 (mature GHPO 1417);

- a 5' oligonucleotide primer having a sequence as shown in SEQ ID NO:54, and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:55 (mature GHPO 1414);

- a 5' oligonucleotide primer comprising a sequence as shown in SEQ ID NO:80 and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:81 (GHPO 750); or

- a 5' oligonucleotide primer comprising a sequence as shown in SEQ ID NO:82 and a 3' oligonucleotide primer having a sequence as shown in SEQ ID NO:79 (mature GHPO 1360).

6. The isolated DNA molecule of claim 5, which can be amplified and/or cloned by the polymerase chain reaction from a *Helicobacter pylori* genome.

7. The isolated polynucleotide of claim 1, which is a DNA molecule that encodes the mature form or a derivative of a polypeptide encoded by the DNA molecule of claim 5.

8. The isolated polynucleotide of claim 1, which is a DNA molecule that encodes the mature form or a derivative of a polypeptide encoded by the DNA molecule of claim 6.

9. A compound, in a substantially purified form, that is the mature form or a derivative of a polypeptide comprising an amino acid sequence that is homologous to an amino acid sequence of a polypeptide associated with the *Helicobacter* membrane, which is selected from the group consisting of the amino acid sequences as shown:

-in SEQ ID NO:2, beginning with amino acid in position -19 and ending with an amino acid in position 689 (GHPO 386);

-in SEQ ID NO:4, beginning with an amino acid in position -20 and ending with an amino acid in position 713 (GHPO 789);

-in SEQ ID NO:6, beginning with an amino acid in position -20 and ending with an amino acid in position 725 (GHPO 1516);

-in SEQ ID NO:8, beginning with an amino acid in position -20 and ending with an amino acid in position 691 (GHPO 1197);

-in SEQ ID NO:10, beginning with an amino acid in position -20 and ending with an amino acid in position 652 (GHPO 1180);

-in SEQ ID NO:12, beginning with an amino acid in position -18 and ending with an amino acid in position 673 (GHPO 896);

-in SEQ ID NO:14, beginning with an amino acid in position -21 and ending with an amino acid in position 619 (GHPO 711);

-in SEQ ID NO:16, beginning with an amino acid in position -17 and ending with an amino acid in position 635 (GHPO 190);

-in SEQ ID NO:18, beginning with an amino acid in position -19 and ending with an amino acid in position 626 (GHPO 185);

-in SEQ ID NO:20, beginning with an amino acid in position -16 and ending with an amino acid in position 467 (GHPO 1417);

-in SEQ ID NO:22, beginning with an amino acid in position -18 and ending with an amino acid in position 673 (GHPO 1414);

- in SEQ ID NO:66, beginning with an amino acid in position -20 and ending with an amino acid in position 279 (GHPO 1360); and

- in SEQ ID NO:68, beginning with an amino acid in position 1 and ending with an amino acid in position 399 (GHPO 750); or

(ii) a derivative of said polypeptide.

10. The compound of claim 9, which is the mature form or a derivative of a polypeptide encoded by a DNA molecule of claim 5.

11. The compound of claim 9, which is the mature form or a derivative of a polypeptide encoded by a DNA molecule of claim 6.

12. A pharmaceutical composition for preventing or treating *Helicobacter* infection in a mammal, said composition comprising a prophylactically or therapeutically effective amount of a compound of claim 9, 10, or 11 and a pharmaceutically acceptable diluent or carrier.

13. The composition of claim 12, further comprising an antibiotic, an antisecretory agent, a bismuth salt, or a combination thereof.

14. The composition of claim 13, wherein said antibiotic is selected from the group consisting of amoxicillin, clarithromycin, tetracycline, metronidizole, and erythromycin.

15. The composition of claim 13, wherein said bismuth salt is selected from the group consisting of bismuth subcitrate and bismuth subsalicylate.

16. The composition of claim 13, wherein said antisecretory agent is a proton pump inhibitor.

17. The composition of claim 16, wherein said proton pump inhibitor is selected from the group consisting of omeprazole, lansoprazole, and pantoprazole.

18. The composition of claim 13, wherein said antisecretory agent is an H_2-receptor antagonist.

19. The composition of claim 18, wherein said H_2-receptor antagonist is selected from the group consisting of ranitidine, cimetidine, famotidine, nizatidine, and roxatidine.

20. The composition of claim 13, wherein said antisecretory agent is a prostaglandin analog.

21. The composition of claim 20, wherein said prostaglandin analog is misoprostil or enprostil.

22. The composition of claim 12, which further comprises a prophylactically or therapeutically effective amount of a second *Helicobacter* polypeptide or a derivative thereof.

23. The composition of claim 22, wherein the second *Helicobacter* polypeptide is a *Helicobacter* urease, a subunit, or a derivative thereof.

24. The composition of claim 12, further comprising an adjuvant.

25. A pharmaceutical composition for preventing or treating *Helicobacter* infection in a mammal, said composition comprising a prophylactically or therapeutically effective amount of a polynucleotide of claim 1, 2, or 3 and a pharmaceutically acceptable carrier or diluent.

26. A pharmaceutical composition for preventing or treating *Helicobacter* infection in a mammal, said composition comprising a prophylactically or therapeutically effective amount of a polynucleotide of claim 5, 6, or 7 and a pharmaceutically acceptable carrier or diluent.

27. A pharmaceutical composition for preventing or treating *Helicobacter* infection in a mammal, said composition comprising a prophylactically or therapeutically effective amount of a polynucleotide of claim 8 and a pharmaceutically acceptable carrier or diluent.

28. A composition comprising a viral vector, in the genome of which is inserted a DNA molecule of claim 4, said DNA molecule being placed under

conditions for expression in a mammalian cell and said viral vector being admixed with a physiologically acceptable diluent or carrier.

29. The composition of claim 28, wherein said viral vector is a poxvirus.

30. A composition that comprises a bacterial vector comprising a DNA molecule of claim 4, said DNA molecule being placed under conditions for expression and said bacterial vector being admixed with a physiologically acceptable diluent or carrier.

31. The composition of claim 30, wherein said vector is selected from the group consisting of *Shigella, Salmonella, Vibrio cholerae, Lactobacillus,* Bacille bilié de Calmette-Guérin, and *Streptococcus*.

32. The composition of claim 25, wherein said polynucleotide is a DNA molecule that is inserted in a plasmid that is unable to replicate and to substantially integrate in a mammalian genome and is placed under conditions for expression in a mammalian cell.

33. An expression cassette comprising a DNA molecule of claim 4, said DNA molecule being placed under conditions for expression in a procaryotic or eucaryotic cell.

34. A process for producing a compound of claim 9, which comprises culturing a procaryotic or eucaryotic cell transformed or transfected with an

expression cassette of claim 33, and recovering said compound from the cell culture.

35. A pharmaceutical composition for preventing or treating *Helicobacter* infection in a mammal, said composition comprising a prophylactically or therapeutically effective amount of an antibody that binds to the compound of claim 9, 10, or 11 and a pharmaceutically acceptable carrier or diluent.

```
1     MKK..L-L.L...L...L.AEDDGFYTSVGYQIGEAAQMV.NTKGIQ.LS GHPO 386
1     MKKHILSLALGSLLVSTLSAEDDGFYTSVGYQIGEAAQMVTNTKGIQQLS GHPO 789
1     MKKHILSLALGSLLVSTLSAEDDGFYTSVGYQIGEAAQMVTNTKGIQQLS GHPO 1516

      MKK..L.L.L...L...L.AEDDGFYTSVGYQIGEAAQMV.NTKGIQ.LS Consensus

50    DNYE.LNNLL..YSTLNTLIKLSADPSAIN..R.NLG.S..NL...K.NS GHPO 386
51    DNYENLNNLLTRYSTLNTLIKLSADPSAINAVRENLGAS.KNLIGDKANS GHPO 789
51    DNYENLNNLLTRYSTLNTLIKLSADPSAINAVRENLGAS.KNLIGDKANS GHPO 1516

      DNYE.LNNLL..YSTLNTLIKLSADPSAIN..R.NLG.S..NL...K.NS Consensus

100   PAYQAVLLA.NAAVGLW.V..YA.T.CG-.G.......G...FNN.PGQD GHPO 386
101   PAYQAVLLAINAAVG.WNV.GY.-T.CG.N.NG.ES.....IFNN.PG.. GHPO 789
101   PAYQAV.LAINAAVGLWN..GYA-..CG-NGNG.ES..G..IFN..PGQD GHPO 1516

      PAYQAV.LA.NAAVG.W....Y....CG.............FN..PG.. Consensus

149   .T.ITCN-....PG.GGP.S..N..K.N.AYQIIQ.AL...------G.N GHPO 386
150   ST.ITC......PG..GPMSI.NFKKLNEAYQILQ.ALKN--G.P.L..N GHPO 789
149   ST.ITCN-.....G.G..MSI..FKKLNEAYQI.Q.ALKN..G.P.LG.N GHPO 1516

      .T.ITC.......G.....S.....K.N.AYQI.Q.AL..........N Consensus

192   G..V.V..N.T......ING.-------K..G.K..T......S....I. GHPO 386
198   ..KVSV.Y.YTC------.G.....C.....G.K..T......S.TT.I. GHPO 789
198   G.KVSV.YNY.C.....ING.....C..K..............TT... GHPO 1516

      ...V.V.............G............................... Consensus

235   TQ...TI.T.----------..NNAQ.LL.QAS.II.TLNEACP.F.. GHPO 386
242   ......I.........T.K.D.....AQ.LL.QAS..I.T.NEACP.F.. GHPO 789
248   TQ...TI.T.......T.K.D-..NNA..LL.QA..I...LN..CP.... GHPO 1516

      ......I...................A..LL.QA.......N..CP.... Consensus

279   -------GG...W.G.S..G..CG.F..EISAIQ.MI.NAQE.VAQSKIV GHPO 386
292   TN..............----.T.G..CG.F..EISAIQ.MI..AQE.V.Q.... GHPO 789
297   TN.....GG...W-G.ST.G..C..F..E.S....MI.NAQE..AQSKIV GHPO 1516

      ..................G..C..F..E.S....MI..AQE...Q.... Consensus

322   SENAQNQN-NLDTGKPFNPYTDASFAQSMLKNAQAQAE.LN.AEQV.KN. GHPO 386
338   ..N.Q........GKPFNP.TDASFAQ.ML.NA.AQA.MLNLA.QV.... GHPO 789
346   SENAQNQN-NLDTGKPFNPYTDASFAQSMLKNAQAQAEM.NL.EQV.KN. GHPO 1516

      ..N.Q........GKPFNP.TDASFAQ.ML.NA.AQA...N...QV.... Consensus
```

Alignment of three predicted polypeptides from H. pylori that share exact identity at their N-terminus (underlined) with the N-terminal amino acid sequence of the mature native 76 kDa protective antigen. A consensus sequence is indicated in bold. The amino acid sequence of GHPO 386 shares 62% identity in a 733 aa overlap with GHPO 789 and 70% identity in a 745 aa overlap with GHPO 1516. Amino acid positions are numbered to the left of the alignment.

Fig. 1 (page 1 of 2)

```
371   E------------------...T.FVS..L..C....G---G...G..PG GHPO 386
388   ----N..N..E..---------.NFV..FLA.C..K.....G...G..PG GHPO 789
395   E...N..N..E..........TNFVS.FLA.C--K.G---G......-- GHPO 1516

      ........................FV...L..C........G........ Consensus

400   .VTSNTWGAGCAYV.QTITNL.NSIAHFGTQEQQIQQAENIADTLVNFKS GHPO 386
425   .VT..T...GCAYV.QT.TNL.NSIAHFGTQEQQIQQAENIADTLVNFKS GHPO 789
438   -VTSNTWGAGCAYV..TI..L.NSIAHFGTQEQQIQQAENIADTLVNFKS GHPO 1516

      .VT..T...GCAYV..T...L.NSIAHFGTQEQQIQQAENIADTLVNFKS Consensus

450   RYSELGNTYNSITTALSKVPNAQSLQNVVSKKNNPYSPQGIETNYYLNQN GHPO 386
475   RYSELGNTYNSITTALSKVPNAQSLQNVVSKKNNPYSPQGIETNYYLNQN GHPO 789
487   RYSELGNTYNSITTALSKVPNAQSLQNVVSKKNNPYSPQGIETNYYLNQN GHPO 1516

      RYSELGNTYNSITTALSKVPNAQSLQNVVSKKNNPYSPQGIETNYYLNQN Consensus

500   SYNQIQTINQELGRNPFRKVGIVNSQTNNGAMNGIGIQVGYKQFFGQKRK GHPO 386
525   SYNQIQTINQELGRNPFRKVGIVNSQTNNGAMNGIGIQVGYKQFFGQKRK GHPO 789
537   SYNQIQTINQELGRNPFRKVGIVNSQTNNGAMNGIGIQVGYKQFFGQKRK GHPO 1516

      SYNQIQTINQELGRNPFRKVGIVNSQTNNGAMNGIGIQVGYKQFFGQKRK Consensus

550   WGARYYGFFDYNHAFIKSSFFNSASDVWTYGFGADALYNFINDKATNFLG GHPO 386
575   WGARYYGFFDYNHAFIKSSFFNSASDVWTYGFGADALYNFINDKATNFLG GHPO 789
587   WGARYYGFFDYNHAFIKSSFFNSASDVWTYGFGADALYNFINDKATNFLG GHPO 1516

      WGARYYGFFDYNHAFIKSSFFNSASDVWTYGFGADALYNFINDKATNFLG Consensus

600   KNNKLS.GLFGGIALAGTSWLNSEYVNLATVNNVYNAKMNVANFQFLFNM GHPO 386
625   KNNKLSLGLFGGIALAGTSWLNSEYVNLATVNNVYNAKMNVANFQFLFNM GHPO 789
637   KNNKLSLGLFGGIALAGTSWLNSEYVNLATVNNVYNAKMNVANFQFLFNM GHPO 1516

      KNNKLS.GLFGGIALAGTSWLNSEYVNLATVNNVYNAKMNVANFQFLFNM Consensus

650   GVRMNLARSKKKGSDHAAQHGIELGLKIPTINTNYYSFMGAELKYRRLYS GHPO 386
675   GVRMNLARSKKKGSDHAAQHGIELGLKIPTINTNYYSFMGAELKYRRLYS GHPO 789
687   GVRMNLARSKKKGSDHAAQHGIELGLKIPTINTNYYSFMGAELKYRRLYS GHPO 1516

      GVRMNLARSKKKGSDHAAQHGIELGLKIPTINTNYYSFMGAELKYRRLYS Consensus

700   VYLNYVFAY GHPO 386
725   VYLNYVFAY GHPO 789
737   VYLNXVFAY GHPO 1516

      VYLNYVFAY Consensus
```

Fig. 1 (page 2 of 2)

```
     MKKLLL-LL----------AEDDGFYTSVGYQIGEAAQMVNTKGI--QL  Consensus

1    MKK.LL-L..........  AED.GF..S.GYQIGEAAQMV...G....L  GHPO 1180
1    MKK.LL-L--........AEDDGFYTSVGYQIGEA.Q.V...G....L  GHPO 896
1    MKK.LL-L--........AEDDGFYTSVGYQIGEA.Q.V...G....L  GHPO 1414
1    MKK..L.L..........-AED.GF..S.GYQIGE.AQMV........L  GHPO 1197
1    MKK.....L..........AED.G...SVGYQIGEA.Q.V........L  GHPO 711
1    MKK..L-L.....-.....AED.GF..S.GYQIGE..QMV...G....L  GHPO 185
1    MKK..L-L.....--....AED.GF..S.GYQIGEA.QMV...G....L  GHPO 190
1    MKK..L----------....AE.DG...SVGYQIGEA.Q.V...G.....  GHPO 1417

     SDNYE-LNNLL---YSTLNTLIKLSADPSAIN----------RNLGSNLKN Consensus

50   SD.YE.L.NLL.....LN........PS.IN...........L.....N  GHPO 1180
48   .D.Y..L.NLL..Y..LN.L..L...PSAI...........NL.S....  GHPO 896
48   .D.Y..L.NLL..Y..LN.L..L...PSAI...........NL.S....  GHPO 1414
50   SD.YE.L.NLL..YS.LNTL...SADP.AIN...........NL....KN  GHPO 1197
51   SD.YE.L..LL.-----N.....S.....IN...........TL.....N  GHPO 711
49   ...YE.L.........L...I.......A.N...........N...N...  GHPO 185
47   ...YE.L...L.....L...I.........N...........N.......  GHPO 190
46   S..YE.L.NLL..Y..L..---..S...S..............L--.....  GHPO 1417

     SPAYQAV-LA-NAAVG-W-----Y----CG-------------FN--PG-  Consensus

100  SPAYQAV.LA.NAAVG.W....-Y...-CG........--.-.F...PG.  GHPO 1180
98   SPAYQAV.LA.NAAVG.W....-....-CG.........---...F...P..  GHPO 896
98   SPAYQAV.LA.NAAVG.W....-....-CG.........---...F...P..  GHPO 1414
100  SPAYQAV.LA.NAA.G.W....-Y....CG............F...P..  GHPO 1197
96   SPAYQA..LA.....G.W....-Y..-.CG............F.......  GHPO 711
99   SP.Y.............W.............................---  GHPO 185
97   SP...AV........G.W...........-.................P..  GHPO 190
92   ..A.QAV..A...AV..W....-....----------------------  GHPO 1417

     ----TITCGS---------------------------------KNAYQ---  Consensus

145  .-...I.C..........----------.............AYQ...  GHPO 1180
144  .-..T.TC...-......----------.............AYQ...  GHPO 896
144  .-..T.TC...-......----------.............AYQ...  GHPO 1414
149  .-..TI.C...-......----------.............AYQ...  GHPO 1197
144  ....TI.CG.....................................AYQ...  GHPO 711
146  -----..C...----------------.....................  GHPO 185
146  -......C...----------------.....................  GHPO 190
118  ----------------------------.................Q...  GHPO 1417
```

Alignment of eight related polypeptides from H. pylori that share significant identity with the consensus sequence determined for the 76 kDa family, a member of which has been determined to be protective in animal models. The amino acid sequence from GHPO 386 shares 53% identity in a 672 aa overlap with GHPO 1180, 51% identity in a 691 aa overlap with GHPO 896, 51% identity in a 691 aa overlap with GHPO 1414, 63% identity in a 711 aa overlap with GHPO 1197, 44% identity in a 640 aa overlap with GHPO 711, 37% identity in a 645 aa overlap with GHPO 185, 36% identity in a 652 aa overlap with GHPO 190 and 41% identity in a 483 aa overlap with GHPO 1417. Amino acid positions are numbered to the left of the alignment. Gaps (-) have been introduced to maximize alignment. Absolute identity shown only (all other residues identified by a dot).

Fig. 2 (page 1 of 4)

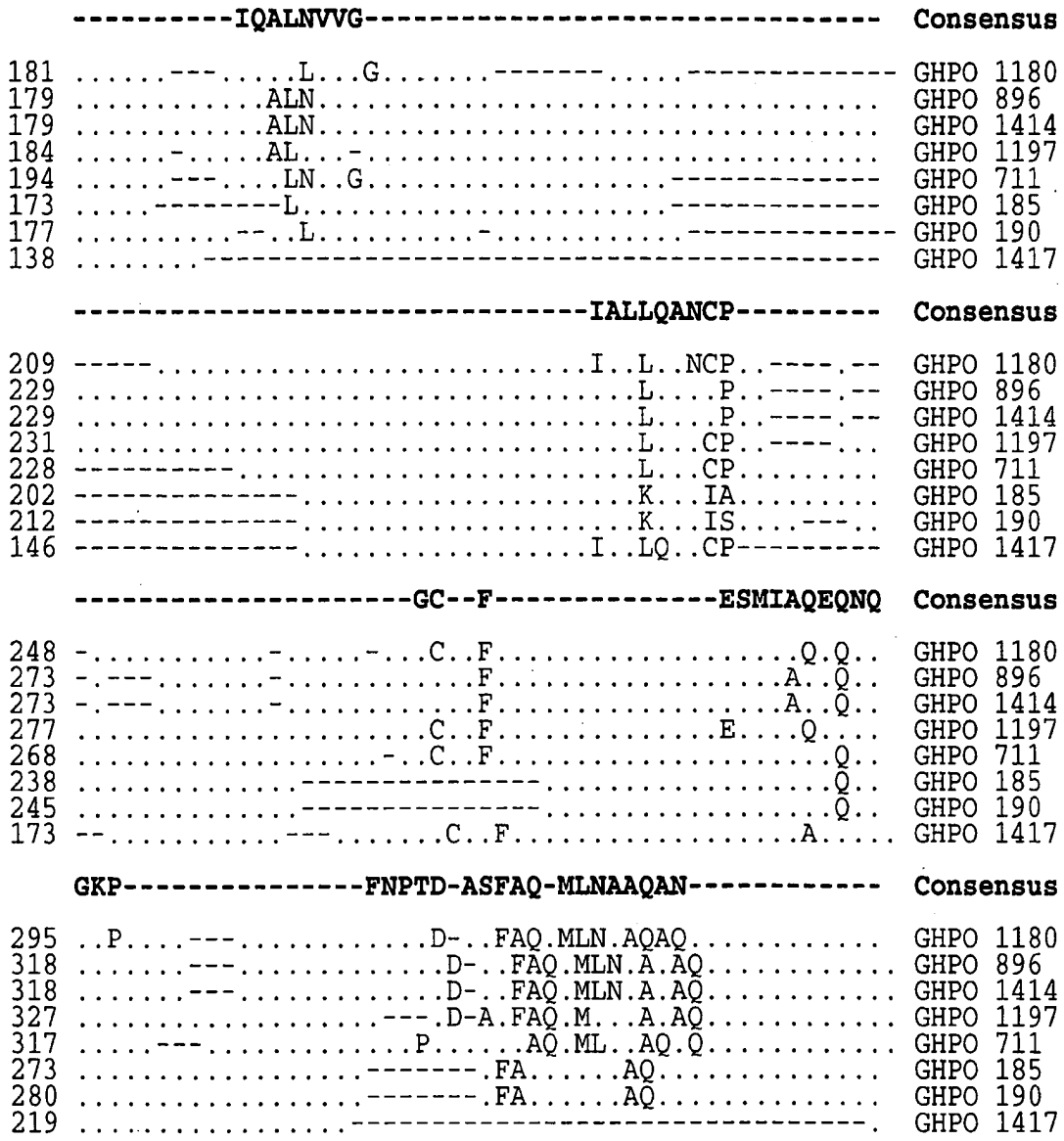

Fig. 2 (page 2 of 4)

```
      ------QVFVL-----C--------------GVTT-----GCAYV----  Consensus

341   ........F.......C.....-----.....GVT......GCAYV....  GHPO 1180
365   ........-.......C.....-----.....T......GCA.V....  GHPO 896
365   ........-.......C.....-----.....T......GCA.V....  GHPO 1414
374   ........FV......C.................VT......GCAYV....  GHPO 1197
365   ..........K.....C.....................---......GCA.V....  GHPO 711
316   ..........K-....--------.........G.T..............  GHPO 185
323   ..........K-....--------.........G.T..............  GHPO 190
237   .......V-------------------------------------.....  GHPO 1417

      ------QVFVL-----C--------------GVTT-----GCAYV----  Consensus

341   ........F.......C.....-----.....GVT......GCAYV....  GHPO 1180
365   ........-.......C.....-----.....T......GCA.V....  GHPO 896
365   ........-.......C.....-----.....T......GCA.V....  GHPO 1414
374   ........FV......C.................VT......GCAYV....  GHPO 1197
365   ..........K.....C.....................---......GCA.V....  GHPO 711
316   ..........K-....--------.........G.T..............  GHPO 185
323   ..........K-....--------.........G.T..............  GHPO 190
237   .......V-------------------------------------.....  GHPO 1417

      -TLN-SIAHFGTQEQQIQQAENIADTLVNFKSRYSELGNTYN---SITTA  Consensus

386   ..LN.S.AHFGTQ..QI.Q.E..A.T...F......L.NTYN---SITT.  GHPO 1180
405   ..L....A..G---.Q..Q....A.T..NFK.....L.....---.I...  GHPO 896
405   ..L....A..G---.Q..Q....A.T..NFK.....L.....---.I...  GHPO 1414
424   ..L..SIAHFG.Q...I..A.N.A.TL.NF...Y..LG.....---SIT.A  GHPO 1197
412   ..L..S.A.F..Q..QI.QA.N.A.TL.----------------------  GHPO 711
357   .T........G---.....A...A....N.KS...E....YN........  GHPO 185
364   .T........G---.....A...A....N.KS...E....YN........  GHPO 190
250   ..L.---------------------------------G.....---.I...  GHPO 1417

      LSKVPNAQ-SLQNVVSKKNNPYSPQGIETNYYLNQNSYNQIQTINQELGR  Consensus

433   .S..PN..-.L.N..S...NP..P.G....Y..NQ..Y.Q.....QELG.  GHPO 1180
449   .S..PNA.-SLQN......NP.SP.G..T-Y.L....YNQ.QTI.QELG.  GHPO 896
449   .S..PNA.-SLQN......NP.SP.G..T-Y.L....YNQ.QTI.QELG.  GHPO 1414
471   LS..P.AQ-SLQNVVSKK.NP.SPQGI..NYY...N...Q.Q...QELG.  GHPO 1197
440   ----------------------------------------------QELG.  GHPO 711
404   .SK.P..Q......V........P.....NY..N.................  GHPO 185
411   .SK.P..Q......V........P.....NY..N.................  GHPO 190
264   L.K.P...-..Q..VS...YS---------Y.LN...Y...QT...E.G.  GHPO 1417
```

Fig. 2 (page 3 of 4)

```
    NPFRKVGIVN-SQTNNGAMNGIGIQVGYKQFFGQKRKWGARYYGFFDYNH  Consensus

482 NPFR.VG...-SQTNNGAMNGIG.Q.GYKQFFG.KR.WG.RYYGFFDYNH  GHPO 1180
497 NPFR..G...-.Q.NNGAMNGIG.QVGYKQFFG.KR.WG.RYYGFFDYNH  GHPO 896
497 NPFR..G...-.Q.NNGAMNGIG.QVGYKQFFG.KR.WG.RYYGFFDYNH  GHPO 1414
520 NPFR..G....S.TNNGAMNGIG.QVGYKQFFG....WGARYYGF.DYNH  GHPO 1197
445 NPFR..G....S.TNNGA.NG.G.QVGYKQFFG.K..WG.RYYGFFDYNH  GHPO 711
454 NPF.KVG....-Q.NNGA.NG.G.QVGYKQFFG....WG.RYYGFFDYNH  GHPO 185
461 NPF.KVG....-Q.NNGA.NG.G.QVGYKQFFG....WG.RYYGFFDYNH  GHPO 190
304 NPFR.VG..N.-Q.NNGAMNG.G.Q.GYKQFFG....FG.RYY.FFDYNH  GHPO 1417

    AFIKSSFFNSASDVWTYGFGADALYNFINDKATNFLGKNNKLS-GLFGGI  Consensus

531 A.IKSSFFNSASDV.TYG.G.D.LYNFINDKAT----KNNK.S.G.FGGI  GHPO 1180
546 A.IKS.FFNSASDVWTYG.G.DALYNFINDK.TNFLGKNNKLS.GLFGG.  GHPO 896
546 A.IKS.FFNSASDVWTYG.G.DALYNFINDK.TNFLGKNNKLS.GLFGG.  GHPO 1414
570 ...KS.FFNS.SDVWTYG.G.D.L.NFINDKAT----K.NK.S.G.FGGI  GHPO 1197
495 A.IKS.FFNSASDVWTYG.G.D.L.NFINDK.TNFLGKNNK.S.G.FGGI  GHPO 711
503 ..IKSSFFNS.SD.WTYG.G.D.L.N.IND..T.---KNNKLS.GLFGGI  GHPO 185
510 ..IKSSFFNS.SD.WTYG.G.D.L.N.IND..T.---KNNKLS.GLFGGI  GHPO 190
353 A.IKS.FFNSAS.V.TYG.G.D.L.NFIN...----.KN.K.S.G.FGGI  GHPO 1417

  ALAGTSWLNSEYVNLATVNNVYNAKMNVANFQFLFNMGVRMNLARSKKKG  Consensus

577 ALAGTSWLNS.YVNLAT.NN.Y.AKMNVANFQFLFN.G.RMNLA..KKK.  GHPO 1180
596 ALAGTSWLNS..VNL...N..YNA.....NFQFLF..G.RMNLAR.KKK.  GHPO 896
596 ALAGTSWLNS..VNL...N..YNA.....NFQFLF..G.RMNLAR.KKK.  GHPO 1414
616 .LAGTSWLNS.YVNLA.VNN.Y.AK.N..NFQFLFN.G.R.NLAR.K..G  GHPO 1197
545 ALAGTSWLNS..VNL.T..NVY.AK.N.ANFQFLFN.G.R.NLAR.KKK.  GHPO 711
550 .LAGT.WLNS.YVNL...NN.Y.AK.N..NFQFLFN.G.R.NLA...KK.  GHPO 185
557 .LAGT.WLNS.YVNL...NN.Y.AK.N..NFQFLFN.G.R.NLA...KK.  GHPO 190
399 ALAGT.WLNS...NL.T.N..Y.AK.N..NFQFLFN.G.R.---------  GHPO 1417

    SDHAAQHGIELGLKIPTINTNYYSFMGAELKYRRLYSVYLNYVFAY      Consensus

627 SDH.AQHG.ELG.KIPTINTNYYS..G..L.YRRLYSVYLNYVFAY      GHPO 1180
646 SDHAAQHGIELG.KIPTINTNYYSFMGA.L.YRR.YS..LNYVFAY      GHPO 896
646 SDHAAQHGIELG.KIPTINTNYYSFMGA.L.YRR.YS..LNYVFAY      GHPO 1414
666 .DH.AQHG.ELG.KIPTINTNYYS..G..L.YRRLYSVYLNYVFAY      GHPO 1197
595 S.HAAQHG.ELG.KIPTINTNYYSF....L.YRRLYSVYLNYVFAY      GHPO 711
600 S.H.AQHGIELG.KIPTI.TNYYSF.G..L.YRRLYSVYLNYVFAY      GHPO 185
607 S.H.AQHGIELG.KIPTI.TNYYSF.G..L.YRRLYSVYLNYVFAY      GHPO 190
440 --....HG.ELG.KIPTINTNYYSFMGA.L.YRRLYSVY.NYV.AY      GHPO 1417
```

Fig. 2 (page 4 of 4)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/06421

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :Please See Extra Sheet.
US CL : 514/ 25, 44, 53, 147; 435/7.32; 424/150.1, 242.1, 451, 653
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 514/ 25, 44, 53, 147; 435/7.32; 424/150.1, 242.1, 451, 653

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P	TOMB, J. et al. The complete genome sequence of the gastric pathogen Helicobacter pylori. Nature, 07 August 1997. Vol. 388, pages 539-547, see entire document.	1-12
Y,P		13-35
X,P	WO 97/12908 A1 (PASTEUR MERIEUX SERUMS AND VACCINS) 10 April 1997, see English abstract, and entire document.	9-12
X	WO 97/11182 A1 (MAX-PLANCK-GESELLSCHAFT ZUR FORDERUNG DER WISSENSCHAFTEN E.V.) 27 March 1997, see English abstract and entire document.	1-12
Y		13-35
Y	US 5,116,821 A (RANDALL et al.) 26 May 1992, see abstract, col. 6, lines 22-68, col. 7, lines 1-41, see entire document.	13-22

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 01 JULY 1998	Date of mailing of the international search report 21 AUG 1998
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer GINNY PORTNER Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/06421

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 5,128,140 A (CHAPURA et al.) 07 July 1992, see abstract, col. 3-5, see entire document.	13-22
Y	US 5,447,923 A (CATRENICH et al.) 05 September 1995, col. 3, 6, and entire document.	13-22
Y,P	US 5,620,964 A (ROTH et al.) 15 April 1997, see abstract, columns 6-8 and entire document.	13-22
Y	US 5,514,660 A (ZOPF et al.) 07 May 1996, column 7-10, see entire document.	13-22
Y,E	US 5,753,630 A (ZOPF et al.) 19 May 1998, column 5, lines 23-31, 7-10, see entire document.	13-22
Y	US 5,476,669 A (BORODY) 19 December 1995, column 1-4, see entire document.	13-22
Y	US 5,610,060 A (WARD et al.) 11 March 1997, column 8, lines 15-41, see entire document.	25-34
A	US 5,567,594 A (CALENOFF) 22 October 1996, see abstract, columns 13-16, and entire document.	1-35
Y	US 5,563,039 A (GOEDDEL et al.) 08 October 1996, see column 2, lines 32-46, column 5, lines 8-14 and entire document.	25-34
X	US 5,258,178 A (CORDLE et al.) 02 November 1993, see abstract, column 4, lines 12-18, 50-58 and entire document.	35
X - Y	US 5,538,729 A (CZINN et al.) 23 July 1996, see column 4, lines 24-25, 44-67 and entire document.	9-12, 35 ---- 30-32
Y	WO 96/40893 A1 (ASTRA AKTIEBOLAG) 19 December 1996, see pages 74-118, and entire document.	1-35
X	FAUCHERE, J. et al. Adherence of Helicobacter pylori cells and their surface components to HeLa cell membranes. Microbial Pathogenesis. 1990, Vol. 9, pages 427-439, see pages 430-431 and entire document.	9-12, 22-23
X	DROUET, E.B. et al. Partial characterization of an external polysaccharide of Helicobacter pylori by using an immunglobulin M monoclonal antibody. Infection and Immunity, June 1993. Vol. 61, No. 6, pages 2732-2736, especially pages 2734-2735.	9-12, 22

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/06421

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	HUSSON, M. et al. Iron acquisition by Helicobacter pylori: Importance of Human Lactoferrin. Infection and Immunity, June 1993, Vol. 61, No. 6, pages 2694-2697, especially page 2697.	9-12, 22
X	ILLINGWORTH, D.S. et al, Siderophore production and iron-regulated envelope proteins of Helicobacter pylori. Zentralblatt fur Bakteriologie, September 1993, Vol. 280, No. (1-2), pages 113-119, especially abstract and page 118.	9-12, 22
Y,P	US 5,679,769 A (DANISHEFSKY et al.) 21 October 1997, column 20, lines 45-65, column 2, lines 26-40.	12-21
A	BOREN, T et al. Attachment of Helicobacter pylori to human gastric epithelium mediated by blood group antigens. Science. 17 December 1993, Vol. 262, pages 1892-1895, especially page 1893.	1-35
A	SHERBURNE, R. et al. Helicobacter pylroi expresses a complex surface carbohydrate, Lewis X. Infection and Immunity. December 1995, pages 4564-4568, especially page 4564-4565.	1-35
A	WADSTROM, T. Microbial adhesion new concepts for development of antiadhesion strategies to combat bacterial infections. ACTA Microbiological Hungary. 22-24 August 1991, Vol. 38, No. (3-4), pages 164-165, especially page 165.	9-24

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/06421

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/06421

A. CLASSIFICATION OF SUBJECT MATTER:
IPC (6):

A01N 43/04, 59/16; A61K 9/48, 31/70, 31/715, 39/02, 39/40; G01N 33/554, 33/569

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

GENEBANK, APS, Dialog
search terms: specific SEQ ID NO., bismuth?, pylori, helicobacter, adjuvant?, urease?, prostagland?, enprostil?, misoprostil?, proton (3a) pump, subcitrate?, subsalicylate?, antisecret?, antibiotic?

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING
This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I, claim(s)1-8,25-34, drawn to no fewer than 84 polynucleotides, vectors containing the polynucleotides, organisms transformed with the polynucleotides, polynucleotides encoding fragments of the polypeptides encoded by the no fewer than 84 different polynucleotides, vaccines and methods of treating a host with the polynucleotides .
Group II, claim(s) 9-24, drawn to polypeptides encoded by the polynucleotides mentioned in Group I.
Group III, claim(s) 35, drawn to a composition comprising antibodies for the treatment of infection.

This application contains claims directed to more than one species of the generic invention. These species are deemed to lack Unity of Invention because they are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for more than one species to be searched, the appropriate additional search fees must be paid. The species are as follows:

Group I contains a separate polynucleotide species for each sequence mentioned. Therefore, there is a minimum of 84 species. Group II contains at least on polypeptide for each polynucleotide sequence mentioned. Therefore, there is a minimum of 19 species in Group II.

For either Group that applicant elects, a total of 10 (ten) specified sequences will be searched and no more than 4 specified sequences will be searched for each additional fee paid; if no additional fee is paid and no election indicated, the first 10 sequences will be searched.

The inventions listed as Groups I, II or III do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:the polypeptide encoding polynucleotides, vectors containing the polynucleotides, organism transformed with said polynucleotides and methods of use are materially different from each other and are therefore independent from the polypeptide of Group II. Additionally, none of the products or methods of Group I is need to make the polypeptide of Group II. The antibodies of Group III are structurally and functional different from the polynucleotides and polypeptides of Groups I and II and therefore, evidence different special technical features which have differing functions and effects. The species listed above do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, the species lack the same or corresponding special technical features for the following reasons: There is no relationship between or among the various nucleotide and amino acid sequences mentioned in the claims.

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☑ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.
As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.